

BAYESED: A GENERAL APPROACH TO FITTING THE SPECTRAL ENERGY DISTRIBUTION OF GALAXIES

YUNKUN HAN^{1,2}, ZHANWEN HAN^{1,2}

Draft version September 1, 2014

ABSTRACT

We present a newly developed version of BayeSED, a general Bayesian approach to the spectral energy distribution (SED) fitting of galaxies. The new BayeSED code has been systematically tested on a mock sample of galaxies. The comparison between estimated and inputted value of the parameters show that BayeSED can recover the physical parameters of galaxies reasonably well. We then applied BayeSED to interpret the SEDs of a large Ks-selected sample of galaxies in the COSMOS/UltraVISTA field with stellar population synthesis models. With the new BayeSED code, a Bayesian model comparison of stellar population synthesis models has been done for the first time. We found that the model by Bruzual & Charlot (2003), statistically speaking, has larger Bayesian evidence than the model by Maraston (2005) for the Ks-selected sample. Besides, while setting the stellar metallicity as a free parameter obviously increases the Bayesian evidence of both models, varying the IMF has a notable effect only on the Maraston (2005) model. Meanwhile, the physical parameters estimated with BayeSED are found to be generally consistent with those obtained with the popular grid-based FAST code, while the former exhibits more natural distributions. Based on the estimated physical parameters of galaxies in the sample, we qualitatively classified the galaxies in the sample into five populations that may represent galaxies at different evolution stages or in different environments. We conclude that BayeSED could be a reliable and powerful tool for investigating the formation and evolution of galaxies from the rich multi-wavelength observations currently available. A binary version of MPI parallelized BayeSED code is publicly available at <https://bitbucket.org/hanyk/bayesed>.

Subject headings: galaxies: fundamental parameters – galaxies: statistics – galaxies: stellar content – galaxies: evolution – methods: data analysis – methods: statistical

1. INTRODUCTION

In the formation and evolution of galaxies, the physical processes about the formation and evolution of stars, the interstellar medium (ISM), and those of super-massive black holes, are expected to be tightly interconnected. The theoretical understanding of these processes and their interactions by using semi-analytical modeling (Cole et al. 2000; Bower et al. 2006; Croton et al. 2006; Baugh 2006; Marulli et al. 2008; Somerville et al. 2008; Bonoli et al. 2009) and numerical simulations (Di Matteo et al. 2005; Springel et al. 2005a,b; Hopkins et al. 2006, 2008a,b; Springel 2010) has been greatly advanced in the last few years. However, due to the huge complexities presented in these physical processes, our understanding of them and their mutual interactions are still far from complete.

Empirical clues from multi-wavelength observations of galaxies (Bell et al. 2003; Shen et al. 2003; Kauffmann et al. 2004; Baldry et al. 2004; Tremonti et al. 2004; Galazzi et al. 2005) are very helpful in providing crucial insights for our further understanding of the formation and evolution of galaxies. Also, the theoretical understanding of the formation and evolution of galaxies needs to be tested, constrained and someday confirmed by many observational results. Fortunately, all of those complex

physical processes involved in the formation and evolution of galaxies have left some imprints in the integrated multi-wavelength spectral energy distributions (SEDs) of galaxies. Therefore, the multi-wavelength SEDs of galaxies are a very important source of information for our understanding of those complex physical processes. The advent of new observing facilities and large surveys at wavelengths from γ -ray to radio (Lonsdale et al. 2003; Jansen et al. 2001; Martin et al. 2005; Giavalisco et al. 2004; Atwood et al. 2009; Abazajian et al. 2009; Scoville et al. 2007; Davis et al. 2007; Driver et al. 2009; Condon et al. 1998) now allow us to obtain the full SEDs of galaxies.

To extract physical information of galaxies from their observed multi-wavelength SEDs, we need some kind of theoretical model for the SEDs of galaxies. The SEDs of most galaxies can be thought of as the superposition of the SEDs of a population of stars with different masses, ages, metallicities, and so on, that constitute the galaxy. Due to this, stellar population synthesis has been the main method of modeling the SEDs of galaxies, starting from the pioneering works of Tinsley (1972), Searle et al. (1973), and Larson & Tinsley (1978). From then on, numerous efforts by different groups have been made to improve this technique (Bruzual 1983; Buzzoni 1989; Bruzual & Charlot 1993, 2003; Bruzual 2007; Leitherer & Heckman 1995; Fioc & Rocca-Volmerange 1997; Maraston 1998, 2005; Zhang et al. 2005; Li & Han 2008; Conroy et al. 2009). However, some important issues still remain. For example, some short lived but bright phases of stellar evolution such as the thermally pulsing asymp-

hanyk@ynao.ac.cn
 zhanwenhan@ynao.ac.cn

¹ Yunnan Observatories, Chinese Academy of Sciences, Kunming, 650011, China

² Key Laboratory for the Structure and Evolution of Celestial Objects, Chinese Academy of Sciences, Kunming, 650011, China

otic giant branch, horizontal branch (Catelan 2009; Lei et al. 2013) and blue stragglers (Tian et al. 2006; Han et al. 2007; Chen & Han 2009) are still not well understood, and they potentially have important effects on the resulting SEDs of galaxies. Furthermore, there are issues about the universality of the stellar initial mass function (Padoan et al. 1997; Myers et al. 2011; Dutton et al. 2013), different parameterizations of the star formation history (Maraston et al. 2010), the complex effects of interstellar medium (Calzetti et al. 2000), the stochastic nature of stellar population modeling (Buzoni 1993; Cerviño 2013), and any possible contribution from active galactic nuclei (Polletta et al. 2007; Murphy et al. 2009; Han & Han 2012). These issues represent the large uncertainties in the modeling of galaxy SEDs, and have resulted in the diversity of SED models. These uncertainties should be properly considered when trying to employ the SED models to derive the physical properties of galaxies, or search for physical relations among these properties (Conroy et al. 2009, 2010; Conroy & Gunn 2010; Conroy 2013).

The main method of extracting physical information from the multi-wavelength SEDs of galaxies is SED fitting. By SED fitting, we try to derive one or several physical properties of galaxies by using certain fitting methods to compare SED models with observed SEDs. In other words, we need to solve the inverse problem: how can the physical properties of galaxies (e.g. stellar ages, stellar masses, star-formation histories, dust extinction and so on) be reasonably derived from quantities that are directly observable (e.g. multi-wavelength photometric SEDs)? In the last decade, the technique of SED fitting has been significantly improved (see, e.g. Walcher et al. 2011, for a recent review). Numerous SED fitting methods and corresponding software have been presented by many authors (Cid Fernandes et al. 2005; Ocvirk et al. 2006; Tojeiro et al. 2007; Koleva et al. 2008, 2009), and have been used by even more authors to derive valuable physical information about the formation and evolution of galaxies.

The modeling of a galaxy’s SED involves the convolutions of the star formation and evolution history, the stellar initial mass function, and the formation and evolution of its dusty interstellar medium, which has nonlinear effects on the resulting SEDs of galaxies, and many other properties that characterize the formation and evolution of the galaxy (Mo et al. 2010). Furthermore, apart from uncertainties in the observations, there are many uncertainties in the modeling of galaxy SEDs. Given these complexities and uncertainties, the problem of deriving the physical properties of galaxies from their directly observable properties is generally not invertible in a strict mathematical sense. However, from the perspective of statistical inversion theory (Kaipio & Somersalo 2004; Tarantola 2005), the inverse problems can be solved by means of Bayesian statistical inference. This is very different from the commonly used approach of solving an optimization problem in the least χ^2 sense (Tinsley & Gunn 1976; Bolzonella et al. 2000; Walcher et al. 2006; Maraston et al. 2006; Koleva et al. 2009; Pforr et al. 2012, 2013; Mitchell et al. 2013; Li et al. 2013), where the main purpose is finding the best fitting result. In Bayesian statistical inference, all quantities (e.g. parameters and fluxes) are modeled as random variables.

Therefore, instead of finding a specific value of a parameter that best matches the observations, the solution to an inverse problem is the posterior probability distribution of the quantity of interest, which describes the degree of confidence about the quantity given the available observations. The posterior probability distributions represent our full knowledge about the parameters, with the degree of uncertainty and degeneracies between them manifesting themselves as easily noticeable broad or multi-peaked distributions. Besides, in Bayesian inference, additional information about the problem can be incorporated as priors to constrain the solution even further. The Bayesian inference methods have been successfully used in many fields of physics (Gregory 2005) and especially cosmology to derive the cosmology parameters of the Universe (Lewis & Bridle 2002; Verde et al. 2003; Cole et al. 2005; Dunkley et al. 2009; Hinshaw et al. 2013).

In recent years, Bayesian methods have been used in the field of SED fitting of galaxies by more and more authors. Benítez (2000) systematically applied Bayesian inference to estimate the photometric redshift of galaxies. Notably, they used Bayesian priors as additional information besides the observed photometric SEDs to give better redshift estimates for the first time. Kauffmann et al. (2003) have used the Bayesian technique to estimate the stellar mass-to-light ratios, dust attenuation corrections and burst mass fractions for a sample of 10^5 galaxies from the Sloan Digital Sky Survey. They also presented a rigorous mathematical description of the method, which currently forms the basis for Bayesian SED fitting. da Cunha et al. (2008) presented an empirical but physically motivated model to interpret the SEDs of galaxies from UV to far-IR consistently, as well as the corresponding MAGPHYS (Multi-wavelength Analysis of Galaxy Physical Properties) package. Similarly, Noll et al. (2009) presented the code CIGALE (Code Investigating GALaxy Emission) for a Bayesian-like analysis of galaxy SEDs from far-UV to far-IR by fitting the attenuated stellar emission and the related dust emission simultaneously.

Recently, the Markov Chain Monte Carlo (MCMC) algorithm has been employed by different authors (Acquaviva et al. 2011; Serra et al. 2011; Johnson et al. 2013) to allow a much more efficient and complete sampling of the parameter space of a SED model than the grid-based methods like CIGALE and MAGPHYS. In Han & Han (2012), we described our BayeSED code, where the multimodal nested (MultiNest) sampling algorithm has been employed, and applied it to a sample of hyperluminous infrared galaxies as a demonstration. The use of MultiNest instead of MCMC allows us to obtain not only the posterior distribution of all model parameters, but also the Bayesian evidence of the model that can be used as a generalization of Occam’s razor for quantitative model comparison. Meanwhile, the principal component analysis and artificial neural networks techniques have been employed to significantly speed up the generation of model SEDs, a major bottleneck for efficient sampling of the parameter space of a SED model.

After the first description in Han & Han (2012), the BayeSED code has been significantly improved, including but not limited to the following. Firstly, the new MultiNest algorithm, which is improved by importance

nested sampling and allows a more efficient exploration of higher-dimensional parameter spaces and more accurate calculation of Bayesian evidence, has been employed. Secondly, besides the artificial neural networks (ANNs) algorithm, the k-nearest neighbors (KNN) algorithm has been added as another method for efficient interpolation of model SED libraries. Thirdly, the redshift of a galaxy can be set as a free parameter, and the effect of the intergalactic medium and Galactic extinction have been considered. Fourthly, the main body of BayeSED has been completely rewritten in C++ in an object-oriented programming fashion, and parallelized with MPI to be able to interpret the SEDs of multiple galaxies simultaneously. In this paper, we systematically test this new version of the BayeSED code with a mock sample of galaxies and apply it to interpret the observed SEDs of a Ks-selected sample of galaxies in the COSMOS/UltraVISTA field with evolutionary population synthesis models.

This paper is organized as follows. In §2, we describe our BayeSED code, and recent improvements to it. We begin in §2.1 by introducing the basic idea of Bayesian inference, and its application to the problem of SED fitting. In §2.2, we introduce the implementation of Bayesian inference with the efficient and robust Bayesian inference tool —MultiNest³, which is capable of calculating the Bayesian evidence of a model and exploring its parameter space, which could be very complex and of a moderately high dimension. To use sampling methods like MCMC and MultiNest, we must be able to evaluate a SED model at any point in the allowed parameter space. So, in §2.3, we present the methods of interpolating a model SED library, while using the evolutionary population synthesis model as an example. In §2.4, we introduce how the MultiNest algorithm and the interpolating algorithm are combined to build up our BayeSED code. To test the ability of BayeSED to recover the physical parameters of galaxies from their multi-wavelength photometry, we employ the method of using a mock sample of galaxies in §3. In §4, we systematically apply our BayeSED code to interpret the SEDs of a Ks-selected sample of galaxies in the COSMOS/UltraVISTA field given by Muzzin et al. (2013). Finally, a summary of our BayeSED method and the results obtained with its application are presented in §5.

2. BAYESED—BAYESIAN SPECTRAL ENERGY DISTRIBUTION FITTING OF GALAXIES

2.1. Bayesian Inference

In BayeSED, we have employed the Bayesian inference methods to interpret the SEDs of galaxies. Bayesian methods have been widely used in astrophysics and cosmology (see, e.g., Trotta 2008, for a recent review). They provide a more consistent conceptual basis for dealing with problems of inference in the presence of uncertainties than traditional statistical methods. For a set of experimental or observational data \mathbf{d} , and a model (or hypothesis) M with some parameters θ that are employed to explain them, the Bayes' theorem states that

$$P(\theta|\mathbf{d}, M) = \frac{P(\mathbf{d}|\theta, M)P(\theta|M)}{P(\mathbf{d}|M)}. \quad (1)$$

For SED fitting, \mathbf{d} represents the observed SED of a galaxy while θ represents the parameters of a SED model M . In Equation 1, $P(\theta|\mathbf{d}, M)$ is called the *posterior probability* of parameters θ given the data \mathbf{d} and model M . $P(\mathbf{d}|\theta, M)$ is the probability of \mathbf{d} given the model M and its parameters θ . It is also known as likelihood $\mathcal{L}(\theta)$, which describes how the degree of plausibility of the parameter θ changes when new data \mathbf{d} is considered. $P(\theta|M)$ is the prior, which describes knowledge about the parameters irrespective of the data. Finally, $P(\mathbf{d}|M)$ is a normalization constant called marginal likelihood, also known as Bayesian evidence.

Bayesian inference is generally divided into two categories: parameter estimation and model comparison. In parameter estimation, the Bayesian evidence $P(\mathbf{d}|M)$, as a normalizing factor that is independent of the parameters θ , is usually ignored. The posterior includes all information that can be used for the complete Bayesian inference of the parameter values. It can be marginalized over each parameter to obtain individual parameter constraints. So, the posterior probability density function (PDF) of a parameter θ_i could be obtained as:

$$P(\theta_i|\mathbf{d}, M) = \int d\theta_1 \cdots d\theta_{i-1} d\theta_{i+1} \cdots d\theta_N P(\theta|\mathbf{d}, M). \quad (2)$$

The Bayesian evidence of a model, which is not important for parameter estimation but critical for model comparison, is given by:

$$P(\mathbf{d}|M) \equiv \int_{\Omega_M} P(\mathbf{d}|\theta, M)P(\theta|M)d\theta, \quad (3)$$

where Ω_M represents the whole N -dimensional parameter space of the model M . It is clear that the Bayesian evidence of a model is just the average of the likelihood weighted by the priors. However, this simple definition automatically implements the principle of Occam's razor: a simpler theory with compact parameter space should be better than a more complicated one, unless the latter is significantly better for the explanation of observational data. Generally, the Bayesian evidence is just larger for a model with a better fit to observations, while smaller for a more complicated model with more free parameters or larger parameter space. The comparison between two models M_2 and M_1 can be formally expressed as the ratio of their respective posterior probabilities given the observational data set \mathbf{d} :

$$\frac{P(M_2|\mathbf{d})}{P(M_1|\mathbf{d})} = \frac{P(\mathbf{d}|M_2)P(M_2)}{P(\mathbf{d}|M_1)P(M_1)}, \quad (4)$$

$P(M_2)/P(M_1)$ is the prior odds ratio of the two models, which is often set to be 1 if none of the two is of special interest. If so, the Bayes factor, which is defined as

$$B_{2,1} \equiv \frac{P(\mathbf{d}|M_2)}{P(\mathbf{d}|M_1)}, \quad (5)$$

can be directly used for Bayesian model comparison. According to the empirically calibrated Jeffrey's scale (Jeffreys 1961; Trotta 2008), $\ln(B_{2,1}) > 0, 1, 1.5$ and 5 (corresponding to the odds of about 1:1, 3:1, 12:1 and 150:1), represent inconclusive, weak, moderate and strong evidence in favor of M_2 , respectively.

³ <http://ccpforge.cse.rl.ac.uk/gf/project/multinest/>

2.2. Sampling of high-dimensional parameter space with MultiNest

Commonly, for the problem of Bayesian parameter estimation, we need to solve the $N - 1$ dimensional integration of Equation 2. However, it is very hard to obtain an accurate analytical solution for this equation in most cases. Moreover, for many problems in astrophysics, we cannot even write down the analytical form of this equation, since the mathematical expression of the priors and/or likelihood function may simply not exist. In practice, Bayesian parameter estimation could be achieved more conveniently by taking a set of samples from the parameter space that are distributed according to the posterior $P(\theta|\mathbf{d}, M)$, where the posterior might be unnormalized. Then, the estimation of parameters could be obtained by some simple statistics of these samples.

The most widely used sampling method for this is the Markov Chain Monte Carlo (MCMC) method. The MCMC technique, which is often based on the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970), provides an efficient way to explore the parameter space of a model and ensures that the number density of samples is asymptotically proportional to the posterior probability density. However, the commonly used MCMC methods are very computationally intensive when the posterior distribution is multimodal or has large degeneracies between parameters, particularly in high dimensions. On the other hand, the calculation of Bayesian evidence, which is critical for Bayesian model comparison, cannot be obtained easily by most MCMC techniques. This is because the evaluation of the multi-dimensional integral in Equation 3 is a challenging numerical task.

The nested sampling, firstly introduced by Skilling (2004), provides an efficient method to calculate the Bayesian evidence, while also produces posterior inferences as a by-product. So, by using the nested sampling method, we are allowed to achieve efficient Bayesian parameter estimation and model comparison simultaneously. This method has been improved further by the works of Mukherjee et al. (2006) and Shaw et al. (2007) to increase the acceptance ratio and the sampling efficiency. Building on these works and pursuing further the notion of detecting and characterizing multiple modes in the posterior from the distribution of nested samples, Feroz & Hobson (2008) introduced the MultiNest algorithm as a viable, general replacement for traditional MCMC sampling techniques. With some further development of this algorithm, the resulting Bayesian inference tool was announced to be publicly released in Feroz et al. (2009b). From then on, the MultiNest algorithm is becoming more and more popular, and has been successfully applied to numerous inference problems in particle physics, cosmology and astrophysics (Trotta et al. 2009; Feroz et al. 2009a, 2010; Martin et al. 2011; Graff et al. 2012; Karpenka et al. 2013; Kavanagh 2014). In Han & Han (2012), we have employed the MultiNest algorithm to build our BayeSED code for SED fitting of galaxies. In the current version of BayeSED, we have employed the most recent version of MultiNest as described in Feroz et al. (2013). The newly developed MultiNest algorithm was largely improved by the technique known as importance nested sampling (INS), which increases the accu-

racy of the calculation of Bayesian evidence by up to an order of magnitude. To achieve the same level of accuracy, the higher evidence accuracy from INS could potentially speed up MultiNest by a factor of a few, if fewer live points or higher target efficiency are used.

2.3. Interpolation of Model SED Library

For the Bayesian inference of the SEDs of galaxies by using an extensive sampling method such as MultiNest (§ 2.2), the SED model needs to be able to be evaluated at any point of its parameter space. However, it would be very computationally expensive to employ a detailed SED model, such as an evolutionary population synthesis model, to generate SEDs during the sampling of a high dimensional and complex parameter space. Besides, for many SED models, only a pre-computed library of model SEDs is available. So, it is often very necessary to interpolate a model SED library. In this subsection, we introduce the interpolation method that we have used in the BayeSED code by taking the evolutionary population synthesis model as an example.

2.3.1. Building of Evolutionary Population Synthesis Model SED Libraries

Currently, the evolutionary population synthesis model is the standard method for modeling the SEDs of galaxies. It is based on the theory of star formation and evolution, the empirical or theoretical stellar spectral library, and the chemical evolution theory of galaxies, and models the SED of a galaxy as the sum of the contribution from individual stars. As mentioned in § 1, there are still many uncertainties in these ingredients of an evolutionary population synthesis model. So, due to different treatments of these issues, there are many competing evolutionary population synthesis models and many possible parameterizations of the model.

In this paper, we have employed the model of Bruzual & Charlot (2003)(bc03) and Maraston (2005)(ma05), two of the most widely used evolutionary population synthesis models. For the bc03 model, the initial mass function (IMF) of Chabrier (2003) and Salpeter (1955) are used, while for the ma05 model, the initial mass function (IMF) of Kroupa (2001) and Salpeter (1955) are used. The star formation histories (SFHs) of galaxies are assumed to be exponentially declining in the form of $\text{SFR} \propto \exp(-t/\tau)$, where t is the time since the onset of star formation and τ is the e-folding star formation timescale. To consider the effect of dust attenuation, a uniform dust screen geometry with a Calzetti et al. (2000) dust extinction law is assumed.

To build SED libraries of the two evolutionary population synthesis models with different assumptions, we have employed a modified version of a grid-based SED fitting code—FAST⁴ (Kriek et al. 2009) to actually generate the model SEDs. We have built four SED libraries, named as ‘bc03_sa’, ‘bc03_ch’, ‘ma05_sa’ and ‘ma05_kr’, respectively. The four SED libraries cover a parameter space with $\log(\tau/\text{yr})$ in the range of [6.5, 11] and in steps of 0.10, $\log(\text{age}/\text{yr})$ in the range of [7.0, 10.1] and in steps of 0.05, and visual attenuation A_V in the range of [0, 4] and in steps of 0.2. For the bc03 model, the

⁴ <http://astro.berkeley.edu/~mariska/FAST.html>

metallicity Z could be 0.004, 0.008, 0.02, or 0.05, while for the ma05 model, Z could be 0.001, 0.01, 0.02, or 0.04. In total, there are 243434 model SEDs for each of the four libraries. A summary of these libraries is presented in Table 1.

2.3.2. Dimension Reduction of Model SED with Principal Component Analysis

A SED model can be considered as a mapping from parameter $\mathbf{X}(x_1, x_2, \dots, x_k)$ to corresponding SED $\mathbf{S}(f_1, f_2, \dots, f_n)$, where x_i represents a parameter and f_i represents a flux at a wavelength. Depending on the resolution of the SED, n could be equal to hundreds or even thousands. So, for a SED model with many free parameters, the size of the required library could be very huge. For the four SED libraries that we have built in §2.3.1, n is just equal to 460, and this results in a size of 1.3Gbyte for each library. However, due to the continuity of a SED, when the flux at a given wavelength is changed, the fluxes at nearby wavelengths will be changed in a very similar way. This means that the fluxes at different wavelengths are not completely independent with each other, and the actual number of dimensions of the SED could be much less than n . So, it is possible to apply some kind of dimensionality reduction technique to efficiently compress a SED.

One such technique is called principal component analysis (PCA), also known as the Karhunen-Loeve transform. It is mathematically defined as an orthogonal linear transformation that transforms the original data to a new coordinate system. The goal of PCA for a SED library is to find an m -dimensional ($m \leq n$) linear model of the n -dimensional SED library that represents the original SEDs as accurately as possible in a least-squares sense. In the current version of our BayeSED code, we have employed the PCA algorithm in the SHARK machine learning library⁵.

We have applied PCA to the four SED libraries that we have built in §2.3.1. It should be noted that we take the logarithm of SEDs before applying PCA. The PCA algorithm in Shark provides two linear models as its outputs. The first linear model, called ‘encoder’, is a linear transformation from an n -dimensional SED \mathbf{S} to an m -dimensional vector $\mathbf{A}(A_1, A_2, \dots, A_m)$, where A_i is the amplitude of the i -th principal component. The second linear model, called ‘decoder’, is the inverse transformation of ‘encoder’. So, it is a linear transformation from an m -dimensional vector \mathbf{A} to an n -dimensional SED \mathbf{S} . The ‘encoder’ is used to compress the SEDs of a library, while the ‘decoder’ is used to reconstruct the SEDs.

The first 3 principal components and mean spectrum of the four SED libraries are shown in Figure 1 and 2. As shown in the figures, the low-order principal components, which determine the general shape of a SED, are more smooth than the high-order principal components, which have more detailed features. For both bc03 and ma05 model, the SED libraries that only differ in the IMF have almost identical principal components, but a slightly different mean spectrum. However, it is worthwhile to notice that the SEDs from bc03 and ma05 models follow different distributions in the space of principal components

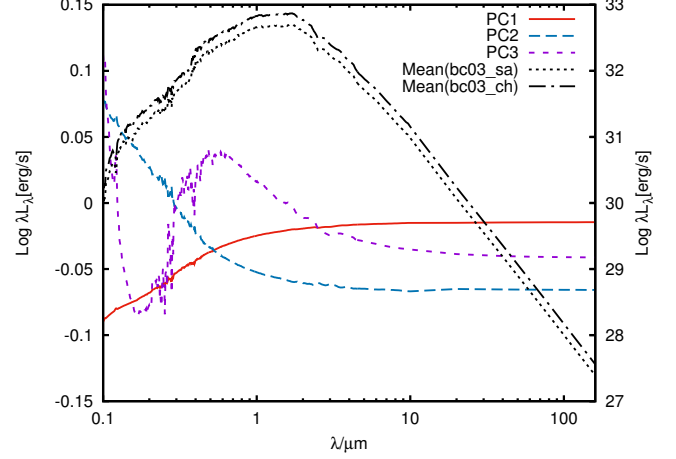


Figure 1. The first 3 principal components (left axis) and mean spectrum (right axis) of the ‘bc03_sa’ and ‘bc03_ch’ model SED libraries. The first 3 principal components contribute 0.82, 0.16, and 0.01 of the overall variance, respectively.

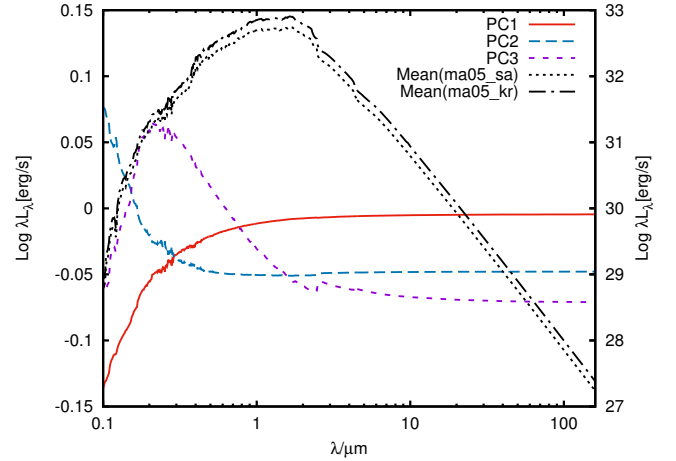


Figure 2. The first 3 principal components (left axis) and mean spectrum (right axis) of the ‘ma05_sa’ and ‘ma05_kr’ model SED libraries. The first 3 principal components contribute 0.80, 0.17, and 0.02 of the overall variance, respectively.

as shown in Figures 3 and 4. While the SEDs from the two models conform to somewhat similar distributions in the PC1-PC2 space, they have very different distributions in the PC2-PC3 space. Meanwhile, the SEDs from the ma05 model show more complex distributions than that from the bc03 model. Furthermore, the relationship between the amplitudes of principal components and physical parameters are different for the two models. Generally, the SEDs from each evolutionary population synthesis model shows a unique distribution in principal component space, a trend which is more obvious in 3D space, as shown in Figure 5. These differences may reflect the consequence of different methodologies and treatments of the TP-AGB phase in the two models.

The total number of principal components m is equal to the dimension of the SED n , which is 460 in our case. However, to compress the original SED libraries, we need to ignore those principal components with much less contribution to the overall variance of the SEDs in the library. In this work, we choose to ignore those principal components with a contribution $\leq 10^{-6}$. This results in 23 principal components for the two SED libraries of

⁵ http://image.diku.dk/shark/sphinx_pages/build/html/index.html

Table 1
Summary of model SED libraries

name	resolution	SFH	IMF	dust extinction law	log(τ /yr)	log(age/yr)	A_V /mag	Z
bc03_sa	pr(460) ^a	exp	Salpeter (1955)	Calzetti et al. (2000)	[6.5, 0.1, 11] ^b	[7.0, 0.05, 10.1]	[0, 0.2, 4]	0.004, 0.008, 0.02, 0.05
bc03_ch	pr(460)	exp	Chabrier (2003)	Calzetti et al. (2000)	[6.5, 0.1, 11]	[7.0, 0.05, 10.1]	[0, 0.2, 4]	0.004, 0.008, 0.02, 0.05
ma05_sa	pr(460)	exp	Salpeter (1955)	Calzetti et al. (2000)	[6.5, 0.1, 11]	[7.0, 0.05, 10.1]	[0, 0.2, 4]	0.001, 0.01, 0.02, 0.04
ma05_kr	pr(460)	exp	Kroupa (2001)	Calzetti et al. (2000)	[6.5, 0.1, 11]	[7.0, 0.05, 10.1]	[0, 0.2, 4]	0.001, 0.01, 0.02, 0.04

Note. — ^a In FAST code, ‘pr’ indicate a photometric resolution SED, and the number of wavelengths for a SED is 460.

^b The parameter is in the range of [6.5, 11] and in steps of 0.1.

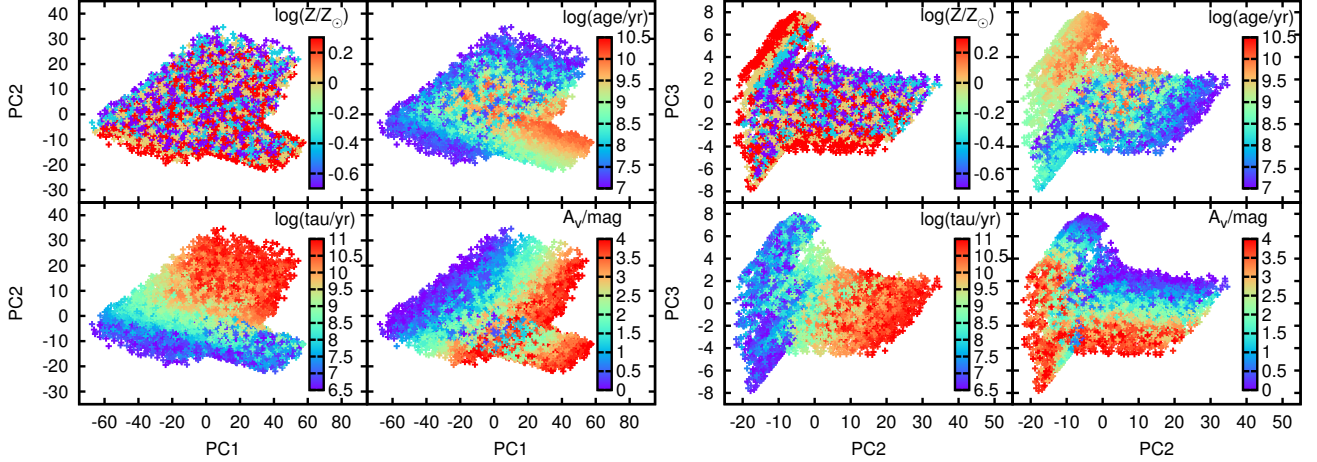


Figure 3. The distribution of model SEDs for ‘bc03_sa’ in the space with PC1-PC2 (left) and PC2-PC3 (right) as basis vectors. The corresponding physical parameters (stellar metallicity, age, e-folding time, and dust extinction) of SEDs are represented by different colors.

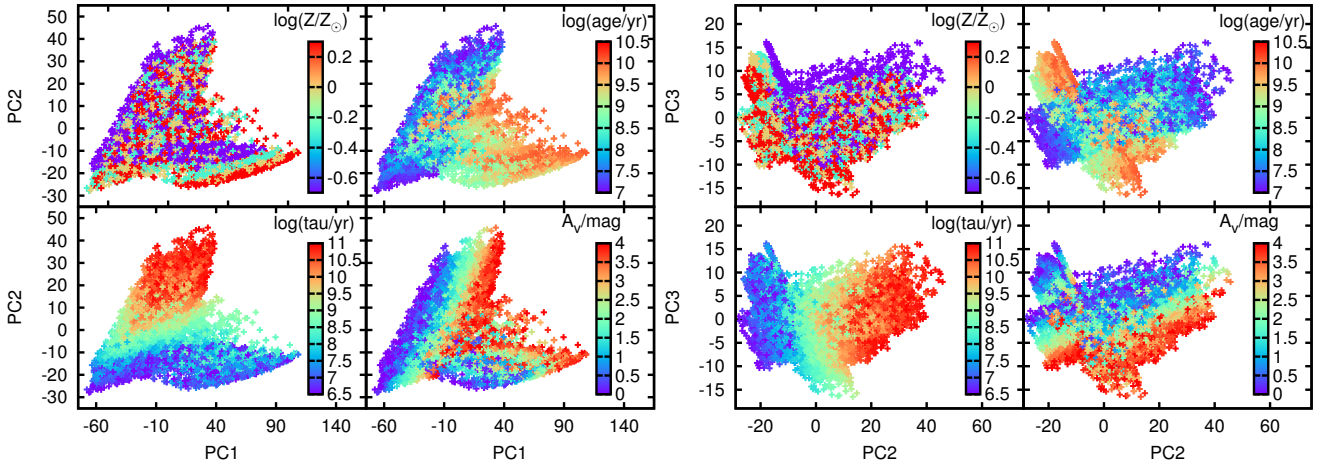


Figure 4. Similar to Figure 3, but for ‘ma05_sa’. The distributions of SEDs for ‘ma05_sa’ are more complex and different from those for ‘bc03_sa’, especially in the PC2-PC3 space.

bc03 model and 26 principal components for the two SED libraries of ma05 model. To check the reliability of the PCA method, we have compared the original SEDs in the ‘bc03_sa’ SED library with those reconstructed from the first 23 principal component. As shown clearly in Figure 6, the reconstructed SEDs are almost identical to the original SEDs in most cases. However, with the application of PCA method, the size of the SED library is reduced to only $23/460 = 5\%$ of the original library.

2.3.3. Interpolation with Artificial Neural Networks

With the application of PCA method, the size of a model SED library is significantly reduced, since a SED can now be represented by the amplitude of only a few

principal components instead of the luminosity at much more wavelengths. On the other hand, the mapping from parameters \mathbf{X} to the corresponding SED \mathbf{S} can now be divided into the mapping from \mathbf{X} to the amplitudes of principal component \mathbf{A} , and the mapping from \mathbf{A} to the final \mathbf{S} . The latter mapping is actually the ‘decoder’ obtained in §2.3.2, which is a linear transformation from \mathbf{A} to \mathbf{S} . So, if we can map \mathbf{X} to \mathbf{A} very quickly, then we would be able to evaluate the original SED model at any point in its parameter space very efficiently. However, as shown in Figures 3, 4, and 5, the relationship between \mathbf{X} and \mathbf{A} could be very complex.

One method of achieving such an efficient mapping from \mathbf{X} to \mathbf{A} is the artificial neural network (ANN) al-

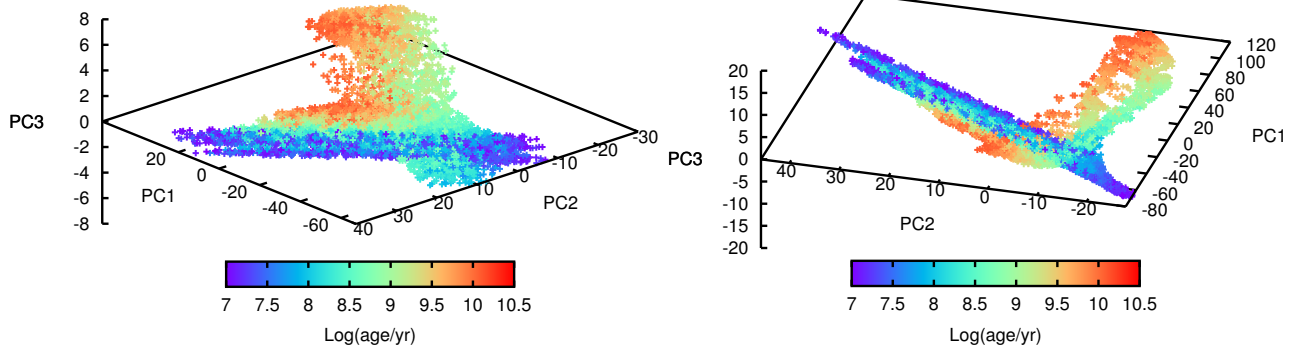


Figure 5. The distribution of model SEDs for ‘bc03.sa’ (left) and ‘ma05.sa’ (right) in the space with the first principal components as basis vectors. The corresponding age of SEDs are represented by different colors. We have selected a point of view that best represents the general shape of the distribution of model SEDs in 3D space.

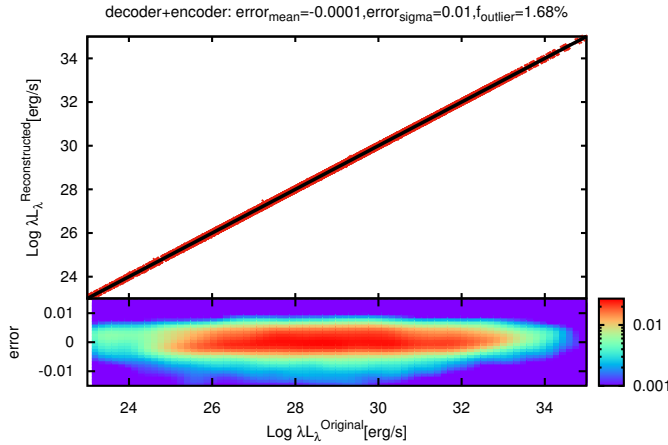


Figure 6. A test of the PCA method by comparing the original SEDs in the ‘bc03.sa’ SED library with those reconstructed from the first 23 principal components. The probability density distribution (in logarithmic scale) of errors is shown in the lower panel. The mean, standard deviation, and the percentage of outliers (error $> 3\sigma$) for the distribution are shown on the top of this figure.

gorithm. ANNs are mathematical constructs originally designed to simulate some intellectual behaviors of the human brain. Just like a human brain, an ANN tries to understand the underlying relationship between two set of things (e.g. \mathbf{X} and \mathbf{A}), from the learning of some instances which obey this relationship. When the procedure of learning is successfully finished, the ANN could be used to predict the corresponding \mathbf{A} from any instance of \mathbf{X} , including those have not been learned before. In the last decade, ANN methods have been successfully used in many problems in astrophysics (Firth et al. 2003; Collister & Lahav 2004; Carballo et al. 2008; Yèche et al. 2010; Almeida et al. 2010; Silva et al. 2011).

Currently, there are many kinds of ANNs and the implements of them using different programming languages. In Han & Han (2012), we have modified the widely used ANNz code (Collister & Lahav 2004), which is originally built for estimating photometric redshifts, to be suit for interpolation of SED models. However, to be able to control every components of an ANN algorithm more freely, we have employed a more general and configurable neural network library—Fast Artificial Neural Network Library⁶—in the current version of BayeSED. As ANNz,

the ANN algorithm implemented in FANN is the most widely used multi-layer perceptron (MLP) feed-forward network. A MLP network consists of a number of layers of neurons. Basically, there are three types layers, which are called input layer, hidden layer, and output layer, respectively. In a feed-forward network, information propagate through the input layer, hidden layers, and output layer sequentially, without any internal feedback. Commonly, the network architecture of such an ANN is denoted as $N_{in}:N_1:N_2:\dots:N_{out}$, where N_{in} is the number of neurons in the input layer, N_i is the number of neurons in i th hidden layer, and N_{out} is the number of neurons in the output layer.

In Figure 7, we show the network architecture of an ANN used for the interpolation of the evolutionary population synthesis model of bc03 and ma05. The input layer of this ANN have 4 neurons, corresponding to the 4 parameters of the evolutionary population synthesis model. These neurons emit the value of corresponding parameters to the neurons in the next layer. It is worth to mention that all parameters have been normalized to have a zero mean and a standard deviation of 1 for a better performance of the ANN. An additional neuron, which is called b_1 , is a bias neuron. The bias neuron play as an offset and always emits 1. The capability of an ANN is mainly determined by the structure of its hidden layers. According to the universal approximation theorem (Cybenko 1989; Kurt & Hornik 1991; Haykin 1999), a multilayer feed-forward network with only one hidden layer can approximate any continuous function to arbitrary precision. However, the neurons in the hidden layer must have a continuous, bounded and nonconstant activation function. Meanwhile, the number of neurons in the hidden layer needs to be chosen carefully according to the complexity of the problem under consideration. As the input layer, an additional bias neuron, b_2 , is also needed. Finally, the last layer gives the output of an ANN. In our case, there is only one output, which is the amplitude of a principal component.

In principle, more neurons in the hidden layer can give better result, but with the expense of much more training time. We practically found that a hidden layer with 30 neurons is good enough for the libraries of bc03 and ma05 model. The choice of activation function for the neurons in the hidden layer and output layer is also crucial. The FANN library give us many possible choices for the activation function. For the neurons in the hidden layer, an

⁶ <http://leenissen.dk/fann/wp/>

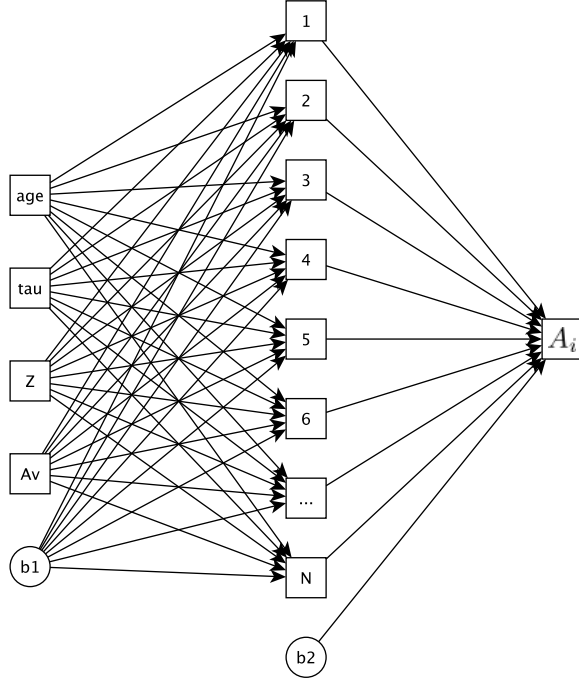


Figure 7. The network architecture of ANN used for the interpolation of the evolutionary population synthesis model of bc03 and ma05. The inputs of this ANN are the 4 parameters of the evolutionary population synthesis model, and the output is the amplitude of a principal component. Here, b1 and b2 are two bias neurons, playing as offsets. The number of neurons in the hidden layer, N, is set to be 30 for both the bc03 and ma05 models. So, the network architecture of ANN used for both bc03 and ma05 models is denoted as 4:30:1.

activation function defined by Elliott (1993) was chosen. This activation function is similar to the commonly used sigmoid activation function, but easier to compute and so faster. For the neuron in the output layer, a linear activation function was chosen to make sure the output could be scaled to any value. Since one ANN gives only the amplitude of one principal component, 23 and 26 of ANNs with the structure as shown in Figure 7 need to be trained for the libraries of bc03 and model, respectively.

The training of an ANN is an optimization problem, where we adjust the weights in the ANN to minimize the difference between the outputs of the ANN and that given by the instances in the training data. The universal approximation theorem of an MLP network states the existence of the solution of such an optimization problem. However, it told us nothing about how to actually find the solution. So, the effectiveness of an ANN method largely depends on the algorithm used for training. The most widely used algorithm for the training of an ANN is the backpropagation algorithm (Rumelhart et al. 1986). As the name suggests, the error obtained by propagating an input through the network is then propagated back through the network while the weights are adjusted in such a way that the error becomes smaller. The original backpropagation algorithm has been further improved by more advanced algorithms like quickprop (Fahlman 1988) and RPROP (Martin & Heinrich 1993; Igel & Hsken 2000). All of these training algorithms have been implemented in the FANN library. As for our cases, we found that the RPROP training algorithm has the best

performance.

Besides the RPROP algorithm, another strategy has been used for the training of ANNs. The set of input and output data, which are parameters of stellar population synthesis model and corresponding amplitudes of the principal components of SEDs in our case, are sorted randomly and then splitted into three groups. The first group of data, called ‘training data’, is used by the RPROP algorithm to adjust the weights in the ANN. During the training of an ANN, the RPROP algorithm tries to minimize the difference between the output of ANN and the results in the ‘training data’. However, if too much training is applied to these data, the ANN will eventually over-fit them. Over-fitting means the ANN will be able to fit the ‘training data’ very precisely, but lose the generalization for other data that have not been used by the RPROP algorithm during the training. So, another group of data, called ‘validation data’, is used to avoid over-fitting. During the training, we trace the difference between the output of ANN and the results in the ‘validation data’, which is called error. The RPROP algorithm will be stopped when the normalized error begin to increase. Finally, the third group of data, called ‘testing data’ will be used for an ultimate test of the ANN when the training is finished.

We have trained 23 and 26 ANNs with the structure as shown in Figure 7 for the libraries of bc03 and ma05 model, respectively. Here, we take the training of 23 ANNs for the ‘bc03_sa’ SED library as an example. The SED library has 243432 SEDs in total. We firstly apply the ‘decoder’, which is obtained by using the PCA method as presented in §2.3.2, to the SEDs in this library to obtain the amplitudes of the first 23 principal components of them. Then, the parameters and corresponding amplitudes of principal components are splitted into ‘training data’ (50%), ‘validation data’ (20%) and ‘testing data’ (30%). These data are used as instances for the training of ANNs. In practice, we actually train the 23 ANNs for the ‘bc03_sa’ SED library simultaneously. Since the code used for ANN training have been parallelized using OpenMP, the whole process of training 23 ANNs for this SED library can be finished in about 40 minutes with 23 Intel 2.20 GHz CPU cores. Finally, all of these ANNs and other information about this SED library are saved to one file, which is only 566 Kbytes in size. This single file will be used to replace the original ‘bc03_sa’ SED library, which is about 1.3 Gbytes in size, when using BayeSED to interpret the observed photometric SEDs of galaxies.

In Figure 8, we give an example of tracing the normalized error of three ANNs on the ‘validation data’ as a function of the number of iteration during the training. As shown in the figure, the normalized error of these ANNs have decreased dramatically during the first few hundreds of iterations. This demonstrated the power of the RPROP algorithm for ANN training. After about 1000 iterations, the decreasing of error becomes very slow, which means the training tend to be converged. As mentioned before, the training will be stopped when the error start to increase. It is worth to notice that the final error for the amplitudes of different principal components are slightly different. It seems that lower order principal component can be ‘learned’ better by the ANN. This is good for us, since the lower order principal

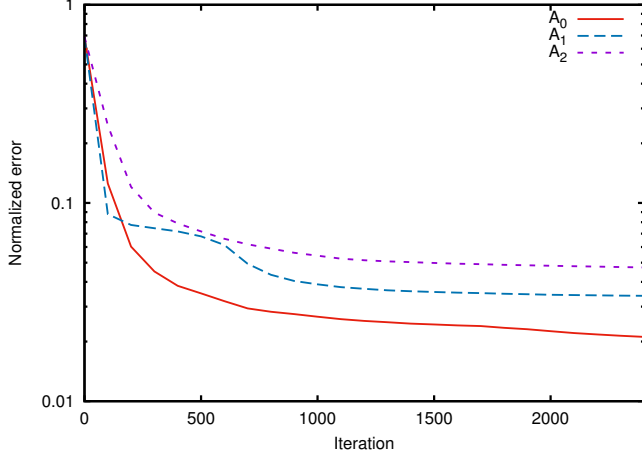


Figure 8. The normalized error on the ‘validation data’ (from ‘bc03.sa’ SED library) as a function of the number of iterations for the training of three ANNs with the structure as shown in Figure 7.

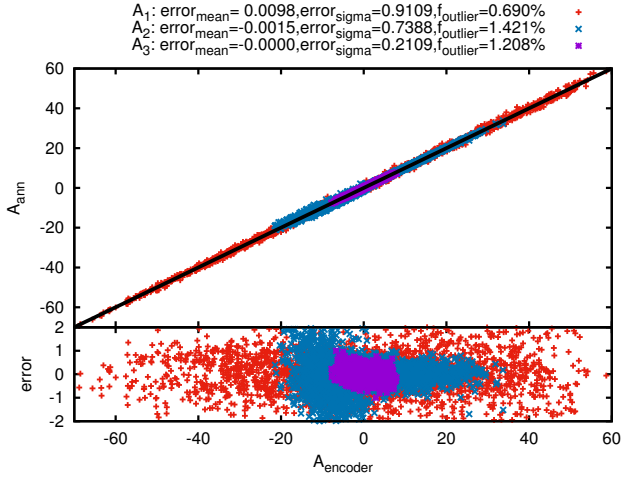


Figure 9. A test of the ANN interpolation method by comparing the amplitudes of principal components of the original SEDs in the ‘bc03.sa’ SED library, which is obtained with ‘encoder’, with that generated with ANN. The results are for the ‘testing data’ set, which have not been used in any way during the training of the ANNs.

component is more important for the reconstruction of SED. We can test the effectiveness of ANN method by comparing the amplitudes of principal components of the original SEDs in the ‘bc03.sa’ SED library, which is obtained with ‘encoder’, with that generated with ANNs. This is shown in Figure 9 for the ‘testing data’ set. As mentioned before, these data have not been used by all means during the training of these ANNs. So, this kind of test should be very rigorous. As shown in the figure, it is clear that the amplitudes of principal components of the original SEDs can be generated pretty well by the ANNs.

Finally, the amplitudes of principal components generated by ANNs can be used to reconstruct the SEDs. This is the ultimate goal of using the ANN interpolation method. In Figure 10, we test the effectiveness of this method by comparing the original SEDs in the ‘bc03.sa’ SED library with that reconstructed by ‘decoder’ from the amplitudes of principal components generated with ANNs. As shown in the figure, the errors

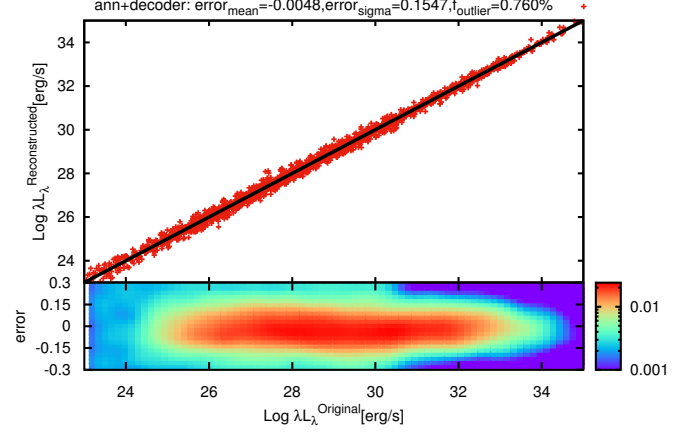


Figure 10. A test of the ANN interpolation method by comparing the original SEDs in the ‘bc03.sa’ SED library with that reconstructed by ‘decoder’ from the amplitudes of principal components generated with ANNs. The probability density distribution of errors is shown in the lower panel. The mean, standard deviation, and the percentage of outliers (error > 3σ) for the distribution are shown on the top of this figure.

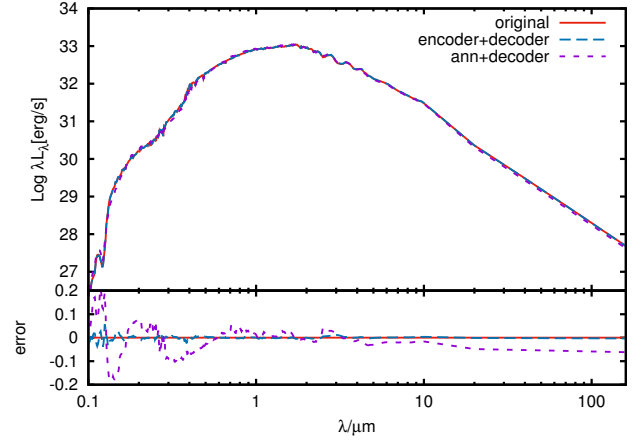


Figure 11. An example of a SED that is reconstructed with the amplitudes of principal components generated with ANNs (ann+decoder), compared with that using the PCA method only (decoder+encoder) and the original one in the ‘bc03.sa’ SED library. The error induced by PCA is negligible, while that induced by ANNs is larger and seems wavelength dependent. However, in most cases, the error is within 0.1 dex.

for SEDs reconstructed by employing ANN method are slightly larger than that obtained with only the PCA method (Figure 6). This is because the ANNs cannot predict the amplitudes of principal components without any error. However, the errors are still very small. In Figure 11, we give an example of SED reconstructed by employing ANN method, and compare it with that obtained using only PCA method and the original one in the ‘bc03.sa’ SED library. It is clear that the original SED can be reconstructed pretty well by the amplitudes of principal components generated with ANNs. So, the method of ANN interpolation of SED library is very successful.

2.3.4. Interpolation with K-Nearest Neighbors

While the ANN method has been proved to be successful for the interpolation of SED library as presented in §2.3.3, there are some reasons for considering other methods for the interpolation of SED library. Firstly,

the network structure of ANN, including the number of hidden layers and the number of neurons in every hidden layer, need to be determined specially for the SED library under consideration. Although an ANN with not too many neurons in one hidden layer has been found to be enough for most problems, there are no simple and general rules for this determination. Secondly, the training of ANNs may need too much time if the relationship between parameters and corresponding SEDs is too complex (highly nonlinear and/or uncontinuous). Thirdly, although the trained ANNs may have a good general performance even for those instances that have not been used during the train, they cannot exactly reproduce those instances that do have been used for the training. One of the methods that can overcome these shortcomings of the ANN method is the interpolation method with K-Nearest Neighbors (KNN). We will use the KNN method as a complement to the ANN method.

The basic idea of the KNN method is very simple. To evaluate at an arbitrary point of the parameter space of an SED model, where only a limited number of results have been given in a SED library, we only need to find the first K nearest neighbors of that point and take the average of the results at these points as the result for that point. The effectiveness of the KNN interpolation method largely relies on how we define the distance between two points, and how we find the first K nearest neighbors. Commonly, the distance between two points is defined as Euclidean distance:

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

To efficiently find the first K nearest neighbors of any point, the knowns points need to be preprocessed into a data structure. Then, a look-up algorithm is applied to this structure to found the nearest neighbors. For this, we have employed a modified version of the Nearest-Neighbor-Regression algorithm in the Shark machine learning library. The library provides two algorithms for the look-up of the K nearest neighbors in a possibly high-dimensional parameter space. They are called ‘SimpleNearestNeighbors’ and ‘TreeNearestNeighbors’, respectively.

The ‘SimpleNearestNeighbors’ algorithm is a brute force algorithm, which just evaluate the distance between pair of points one by one. So, the organization of known points is not important for this algorithm. The ‘TreeNearestNeighbors’ algorithm, nevertheless, need the known points to be organized as some kind of tree structure in advance. Generally, the ‘TreeNearestNeighbors’ algorithm is much faster than the ‘SimpleNearestNeighbors’ algorithm if only the dimension of the data points is not too high. The Shark machine learning library provides tree choices for the tree structure, named as KDTree, KHCTree, and LCTree, respectively. They all belong to the binary space-partitioning tree. The KDTree, standing for k-dimensional tree (Friedman et al. 1977), is the most widely used algorithm for nearest-neighbor search. It works well in low-dimensional data, but quickly loses its effectiveness as dimensionality increases. The KHCTree and LCTree, standing for Kernel Hierarchical Clustering tree and Linear Cut tree, respec-

tively, are more advanced tree structures (see the document of Shark library ^{7,8} for more details about the two tree structures). We practically found that the LCTree structure has the best performance for our case.

As the ANN interpolation method in §2.3.3, we have applied the KNN interpolation method to predict the amplitudes of principal components. It is possible for us to predict the corresponding SED directly from the given value of input parameters. However, for KNN interpolation, we need to save all instances of a SED library to a file in the disk, and reload it into the memory of computer during the sampling of parameter space. It is clear that the applicability of this method is largely limited by the size of the SED library. As shown in §2.3.2, the size of a SED library can be reduced to only 5% of the original by using PCA method. So, by combined with the PCA method, the KNN interpolation method could be more useful in practice. In Figures 12 and 13, we show a test of KNN interpolation method, which is similar to that for ANN method in Figures 9 and 10. We should mention that the instances used for this test has not been used to build the LCTree that are use for interpolation. So, as the case for the test of ANN method using ‘testing data’, this should be a rigorous test. It is clear that for both the amplitudes of principal components and the reconstructed SEDs, the KNN interpolation method could be even better than the ANN interpolation method.

There is no doubt that the KNN interpolation method has some advantages over ANN interpolation method. For KNN interpolation, an intensive training process, which is crucial for ANN, is unnecessary. We only need to store the known instances of a model properly in a data structure that is convenient for searching. However, this does not mean that KNN method is better than ANN method in all aspects. For example, the ANN interpolation is much faster than KNN interpolation during the sampling of parameter space. Besides, the size of data that need to be store for KNN interpolation (e.g., 46 Mbyte for the ‘bc03_sa’ SED library) is much larger than that for ANN (566 Kbytes). Furthermore, the KNN method is much more sensitive to outliers and local structure of the data than ANN method. So, in practice we use KNN method as a complement to the ANN method. Then, it would be worth to check whether the results could be different by using the two methods.

2.4. The building-up of BayeSED code

As the last part of this section, we introduce how the MultiNest (§2.2) and ANN (§2.3.3) or KNN (§2.3.4) algorithm are combined to build up our BayeSED code for interpreting the multi-wavelength SEDs of galaxies.

In Figure 14, the flowchart of BayeSED code is shown. The main inputs of BayeSED are the observed multi-wavelength photometric SED, including measurement errors, of a galaxy that needs to be interpreted. On the other hand, the priors for the SED model that is used to explain the observations are considered as additional inputs. This includes the allowed ranges for all free parameters of the model, and corresponding distributions

⁷ http://image.diku.dk/shark/doxygen_pages/html/classshark_1_1_k_h_c_tree.html#details

⁸ http://image.diku.dk/shark/doxygen_pages/html/classshark_1_1_l_c_tree.html#details

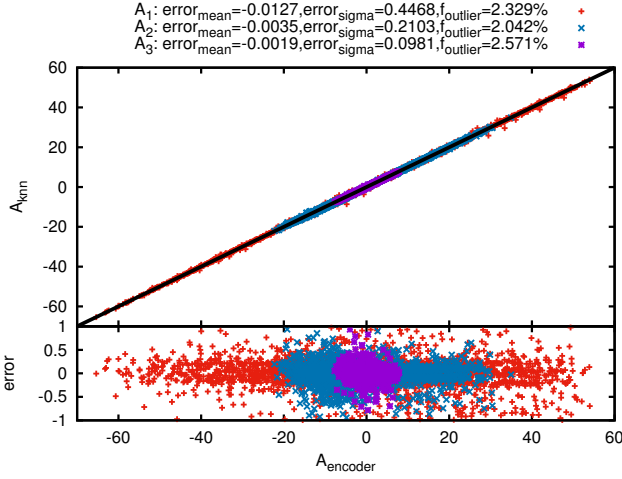


Figure 12. A test of KNN interpolation by comparing the amplitudes of principal components of the original SEDs in the ‘bc03.sa’ SED library, which is obtained with ‘encoder’, with that generated with KNN.

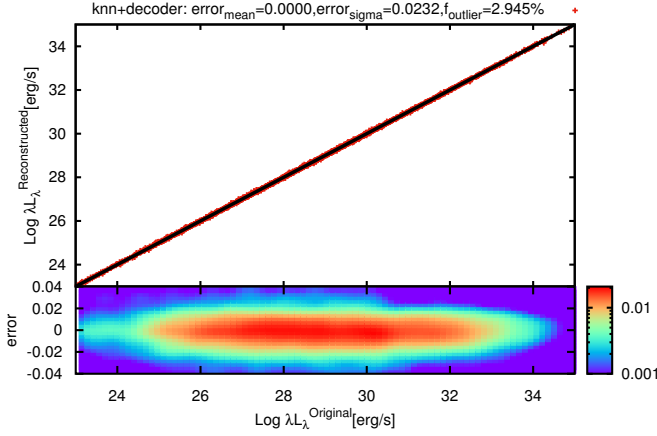


Figure 13. A test of KNN interpolation by comparing the original SEDs in the ‘bc03.sa’ SED library with that reconstructed by ‘decoder’ from the amplitudes of principal components obtained with KNN. The error induced by KNN interpolation is much smaller than that induced by ANN interpolation (Figure 10), and only slightly larger than that induced by PCA alone (Figure 6).

of them. Currently, the distribution of a free parameter is only allowed to be uniform. So, for parameters with a large dynamic range, the logarithm of them should be used as free parameters instead. In the future, we plan to provide more choices for the distribution of priors, including those that are physically more informative.

The sampling with MultiNest of the parameter space of a SED model that is used to explain the observations lies at the heart of BayeSED. During the sampling, the MultiNest sampler continuously request that the likelihood function at a specific point of the parameter space should be computed, until the resulting posterior and Bayesian evidence are thought to be converged. The computation of likelihood at a give point involves the computation of the model SED at that point. This is achieved by using the ANN algorithm (§2.3.3) or KNN algorithm (§2.3.4) to interpolate a pre-computed SED library. As mentioned before, the computation of model SED is a major bottleneck for an efficient sampling of the possibly high-dimensional and complex parameter space of a SED model. Thanks to the ANN and KNN algo-

rithm, this can be done very quickly in our BayeSED code. The huge SED libraries are only used to build the ANNs or KNNs once and for all. So, they are no longer needed during the sampling of parameter space, while a small file including all necessary information is used instead.

When a model SED is generated with ANNs or KNNs, the effects of cosmological redshift and intergalactic medium (IGM) on the SED are further considered. In the current version of BayeSED, we provide two options for considering the effects of IGM. They are based on the prescription of [Madau et al. \(1996\)](#) and [Meiksin \(2006\)](#), respectively. Optionally, the effect of Galactic dust reddening and extinction could be considered by setting the value of $E(B - V)$ at the position of the object, where the R-dependent Galactic extinction curve of [Fitzpatrick \(1999\)](#) with the ratio of total to selective extinction $R(V) = A(V)/E(B - V) = 3.1$ is used. The redshift of galaxy is considered as an optional free parameter for the fitting of SED. Then, the redshift and other physical parameters of a galaxy can be obtained simultaneously with our BayeSED code. However, BayeSED is not optimized for the determination of redshift, while many publicly available codes have been designed for that purpose. So, the redshift of galaxies could be determined by using other codes and then used in BayeSED. We will test the reliability of using BayeSED to determine the redshift of galaxies in §4. Finally, the model SED is convolved with the transmission function of filters to obtain model fluxes that are directly comparable with multi-wavelength observations.

The value of likelihood at a specific point of the parameter space as requested by the MultiNest sampler is obtained by the comparison between the model fluxes and the corresponding multi-wavelength observations. Commonly, the distribution of observational errors are assumed to be Gaussian. Then, the normalized likelihood function is defined as:

$$\mathcal{L}(\theta) = \prod_{i=1}^{i=n} \frac{1}{\sqrt{2\pi}\sigma_{o,i}} \exp\left(-\frac{1}{2} \frac{(F_{o,i} - F_{m(\theta),i})^2}{\sigma_{o,i}^2}\right), \quad (7)$$

where $\sigma_{o,i}$ is the observational error in the i -th band, $F_{o,i}$ and $F_{m(\theta),i}$ are the observed flux and model flux in the i -th band, respectively. In practice, the term $1/\sqrt{2\pi}\sigma_{o,i}$ is usually omitted, since it is independent of the shape of the likelihood function. So, in most works of Bayesian SED fitting, the definition of likelihood function is simplified as:

$$\mathcal{L}(\theta) = \prod_{i=1}^{i=n} \exp\left(-\frac{1}{2} \frac{(F_{o,i} - F_{m(\theta),i})^2}{\sigma_{o,i}^2}\right). \quad (8)$$

In the above definition of likelihood, the observations at different wavelength bands are assumed to be independent and only observational error have been considered. The possible systematic error of the SED model, which could be important, especially for population synthesis models ([Conroy et al. 2009](#); [Cerviño 2013](#)), have not been considered yet. The systematic error of a SED model is likely wavelength and model dependent, and so not easy to be considered properly. In the EAZY code ([Brammer et al. 2008](#)), this has been considered as a template error

function. However, it is not clear how universal this kind of error function could be.

On the other hand, thanks to the application of PCA, an SED can be described by the amplitudes of principal components. Then, the likelihood function could be defined as:

$$\mathcal{L}(\theta) = \prod_{i=1}^{i=N} \exp\left(-\frac{1}{2} \frac{(A_{o,i} - A_{m(\theta),i})^2}{\sigma_{o,i}^2}\right), \quad (9)$$

where N is number of principal components. In this approach, the model SEDs need not to be reconstructed from the amplitudes of principal component over and over again, while the observed SED needs to be projected to the principal components only once. This is especially useful for the analysis of spectroscopic data (Chen et al. 2012), where the dimension of the data is much larger than the number of necessary principal components. However, this is less helpful for the analysis of photometric data, where the dimension of the data is comparable to the number of necessary principal components. Besides, it is not so straightforward to project the sparsely sampled photometric SED to the principal components (see also Wild et al. 2014).

By using the BayeSED code to interpret the multi-wavelength SED of a galaxy, many outputs could be obtained. Firstly, we can obtain the Bayesian evidence of the model which is used to explain the observed SED. Secondly, we can obtain the estimation of all parameters of the model. As mentioned in §2.1, the posterior PDF including all information about the parameters. However, in practice, it is not possible to report the results of Bayesian parameter estimation with full PDF, especially for a large sample of galaxies. So, it is very necessary to use some summary statistics instead. In BayeSED, we provide many summary statistics about a parameter, including mean, median, maximum-a-likelihood (MAL, or best-fit), and maximum-a-posteriori (MAP). The corresponding error of a parameter is estimated with standard deviation or percentiles of the PDFs. Besides, the best-fit model SED and corresponding amplitudes of principal components, rest-frame absolute magnitudes, observed-frame apparent magnitudes could be optionally outputted. Finally, it is worth to mention that BayeSED has been fully parallelized with MPI⁹. So, it is possible to interpret the multi-wavelength SEDs of a large sample of galaxies simultaneously, and all results are saved into a single file.

3. APPLICATION TO A MOCK SAMPLE OF GALAXIES

As mentioned in §1, by interpreting the SEDs of galaxies, we try to solve the inverse problem of SED modeling: to derive the physical parameters of galaxies from their observed multi-wavelength photometric SEDs. The ability of a SED-fitting code to solve this problem can be properly tested by using mock samples of galaxies. So, before applying the BayeSED code to interpret the SEDs of galaxies in a real sample, we will test the reliability of it in this section.

⁹ The MPI implemented in the MultiNest algorithm itself is switched off, since we found that it is not efficient for multiple SED-fitting.

3.1. Building of Mock Sample of Galaxies

The starting point for the building of a mock sample of galaxies are a set of model SEDs of galaxies, which are obtained with a model of galaxy SED such as the evolutionary population synthesis model. Then, these model SEDs are transformed according to that would be experienced of real galaxies to obtain mock fluxes at multiple wavelength bands. Since the true value of all parameters of the mock observations are known in advance, it is easy to check if they can be recovered properly by a SED-fitting code.

In our case, we started from the SED libraries as built up in §2.3.1 to make four mock samples of galaxies. We have taken the bc03 model with a Salpeter (1955) IMF as an example. From the SED library, a total of 10000 model SEDs are randomly selected to make a mock sample with 10000 galaxies. These model SEDs are shifted to a random redshift z ranging from 0 to 6, while also consider the effect of IGM. We demand that the age of a galaxy must be smaller than the age of Universe at that redshift. Then, the model SEDs are convolved with the transmission function of filters to obtain model fluxes. Finally, some random noises with a Gaussian distribution are added to these model fluxes. The Gaussian distribution has a zero mean and a dispersion equal to 10% (S/N) of the model flux. The filters and corresponding errors are selected to mimic the Ks-selected sample of galaxies in COSMOS/UltraVISTA field that will be studied in the next section.

It should be mentioned that there are other methods to build more realistic mock sample of galaxies. For example, the distribution of luminosity and redshift could be drawn from a luminosity function of galaxies, or the distributions of physical parameters of galaxies could be that predicted by a model for the formation and evolution of galaxies. However, these more realistic mock samples are not very necessary for a reasonable test of a SED-fitting code. A good SED-fitting code should be able to properly recover the physical parameters of galaxies from their multi-wavelength observations, regardless of how these parameters are distributed. Nevertheless, we should keep in mind that not all physical parameters of galaxies can be recovered equally well, even though the best possible SED-fitting code has been used. Besides the code itself, there are many other factors that can take part in determining the possibility of recovery. For example, the number of available filters and corresponding S/N, the relative importance of a parameter for determining the shape of SED, the degeneracies between parameters, and so on. So, the mock sample of galaxies is only used to check the internal consistency of BayeSED and the effects of intrinsic degeneracies between parameters of a SED model (see e.g. Walcher et al. 2008).

3.2. Interpreting and Results

We have applied our BayeSED code to interpret the mock sample of galaxies that are built up in §3.1. For the interpolation of model SEDs, both ANN and KNN methods have been used, and the results obtained will be compared here. It is worth to mention that the results presented here represent an overall verification of everything that are involved in the BayeSED code. Since the

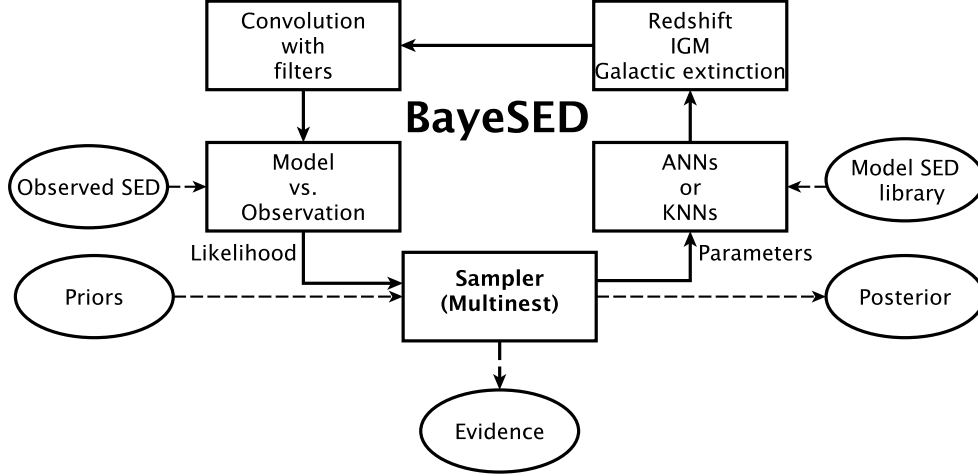


Figure 14. The flowchart for interpreting the SED of galaxies with BayeSED. Since BayeSED has been fully parallelized with MPI, this kind of analysis could be done for multiple galaxies, simultaneously.

mock sample is built up with the original SED library, all potential errors that are hidden in the programming of code, or the PCA, ANN, KNN and MultiNest algorithms could propagate into the final results.

Since the ANN and KNN methods are used to approximate the original bc03 model, they can be considered as two special versions of that model. So, it is meaningful to check the Bayesian evidences of them for the mock observations that are built up from the original bc03 model. In Figure 15, we show the probability density distribution function (PDF)¹⁰ and cumulative distribution function (CDF) of Bayes factor $\ln(B_{\text{knn,ann}})$ for the mock sample of galaxies. It is clear from the figure that the results obtained with KNN method has larger Bayesian evidence than that obtained with ANN method. Except for the method of the interpolation, anything else is the same for the two approaches. So, this indicate that KNN method should be a better approximation to the original bc03 model.

In Figures 16, 17, 18, 19, and 20, we compare the recovered values of redshift, age, metallicity, e-folding time, and dust extinction with their true values. Among these, the redshift can be best recovered. For both the ANN and KNN methods, the mean for the distribution of errors are almost zero, while the dispersion and the fraction of outliers with an error larger than 3σ is also very small. The age and dust extinction of galaxies can be recovered moderately well. This can be better understood from the posterior probability density functions (PDFs) of them for a mock galaxy in Figure 21. The PDFs of $\log(\text{age/yr})$ and A_V/mag show a weak second peak, indicating the degeneracy between them. Finally, the metallicity and e-folding time cannot be recovered properly. This seems because these two parameters are less important for shaping the SED, and so more easily be affected by errors in the observations and the degeneracies with other parameters. As can be noticed in Figure 21, the PDFs of $\log(Z/Z_\odot)$ and $\log(\text{tau/yr})$ show many peaks, indicating serious degeneracies with other parameters. Meanwhile, it is clear from Figures

18 and 19 that the distribution of errors for them show a clear anti-correlation with the true value. This kind of trends just imply that the two parameters can not be constrained effectively by the observations, but mainly constrained by the allowed range of them. Since a flat prior is assumed and the parameters are estimated with the mean of PDF, the parameters tend to be overestimate at the lower end and underestimated at the higher end. In practice, we found that if the redshift is fixed to the right value, the two parameters can be recovered much better. However, it seems still difficult to recovery them properly with photometric data only.

With the estimated value of free parameters of an evolutionary population model, we are able to derive other parameters, such as the stellar mass and star formation rate. However, it would be more convenient to be able to estimate the free parameters and other derived parameters simultaneously. This is allowed in our BayeSED code. We achieved this by building another set of ANNs or KNNs to derive other parameters from the free parameters of the SED model. This is similar to that for SED, except that the output of ANNs or KNNs are the derived parameters instead of SED. With this method, we are able to simultaneously estimate any number of parameters that are derived from the free parameters of the SED model.

As examples, in Figures 23, 24, and 25, we compare the derived values of stellar mass, stellar bolometric luminosity, and star formation rate with their true values. Generally, these parameters can be recovered properly except for a small fraction of extreme outliers. Besides, the recovered value of these parameters seem to be biased for the outliers. This is especially clear for the estimation of star formation rate. We found that this bias is mainly depends on the method used for summary statistics. In Figure 26, we show the results obtained with MAL estimation. We can see that the distribution of errors is more symmetric with MAL estimation, and the mean of errors is much closer to zero. However, the dispersion and the fraction of outliers are not changed too much. So, for a population of galaxies, the Mean estimation and MAL estimation should be equally good.

In this section, we have systematically tested the reli-

¹⁰ The PDFs hereafter in this paper are obtained with the method of kernel density estimation.

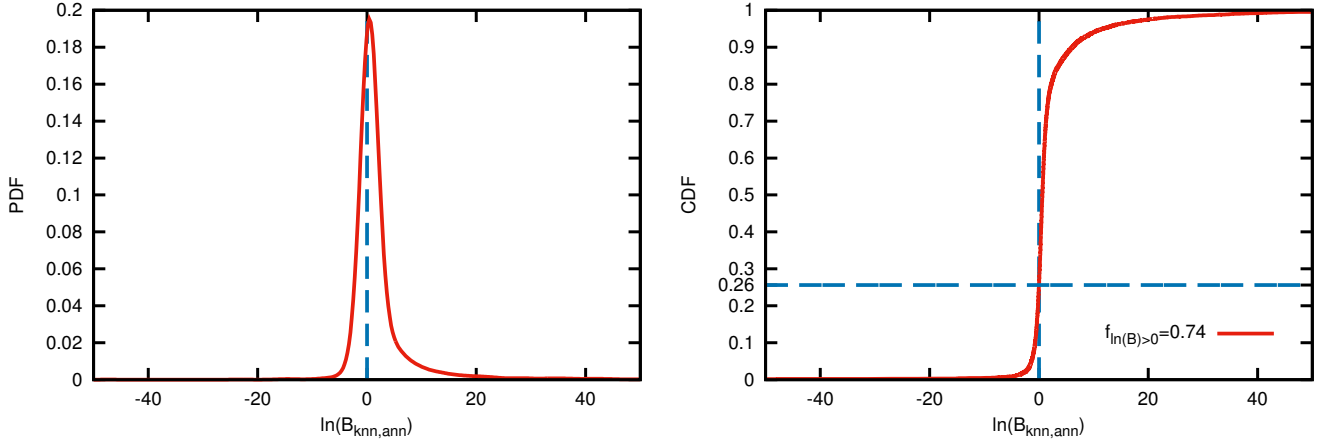


Figure 15. The probability density distribution function (PDF) and cumulative distribution function (CDF) of the Bayes factor $\ln(B_{\text{knn,ann}})$ for the mock sample of galaxies. The bc03 model with a [Salpeter \(1955\)](#) IMF was used in this example. The KNN method gives a much better approximation to the original bc03 model than the ANN method, since only 26% of the mock galaxies favor the latter.

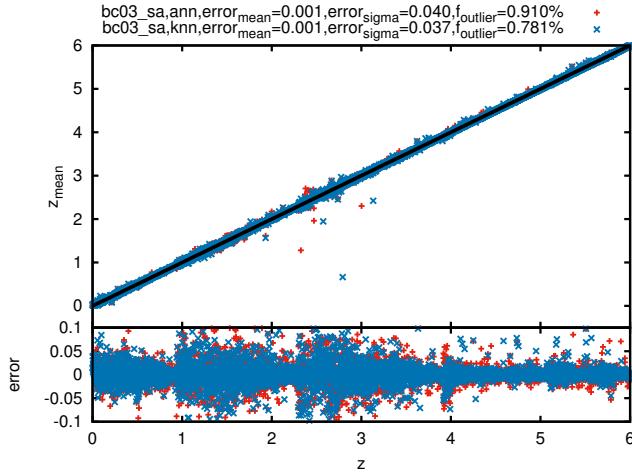


Figure 16. The recovered values of redshift (z_{mean}) compared with the true values (z) of the mock sample of galaxies. The redshifts and corresponding errors are estimated from the mean and standard deviation of the posterior distribution. The bc03 model with a [Salpeter \(1955\)](#) IMF was used in this example. On the top of this figure, the mean, dispersion and the fraction of outliers with an error larger than 3σ of the distribution of errors are shown.

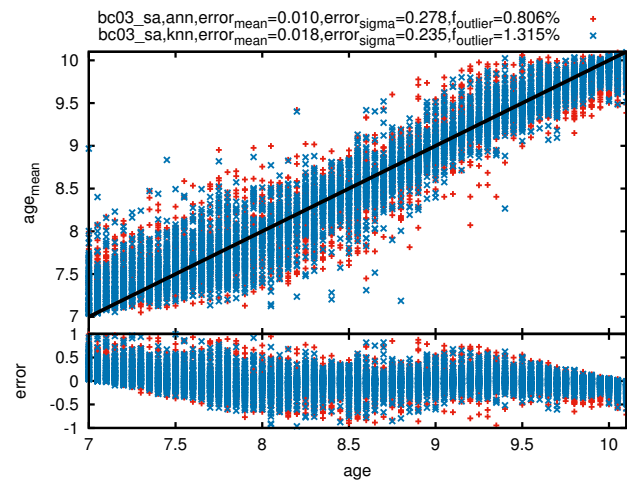


Figure 17. Similar to Figure 16, but for the age of galaxies. The age tends to be overestimated at the lower end while underestimated at the higher end. This is mainly caused by the limited parameter range.

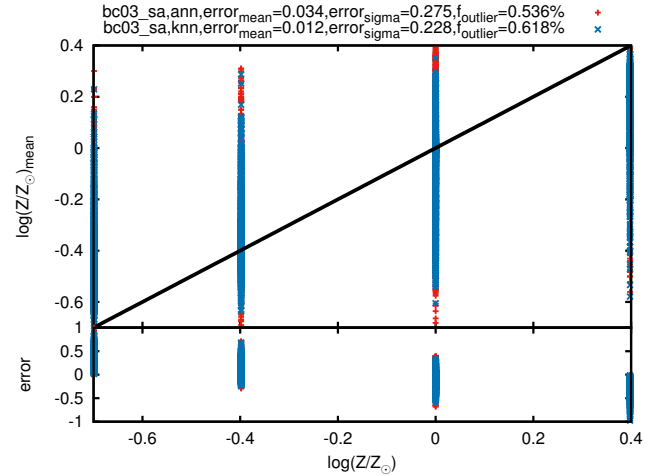


Figure 18. Similar to Figure 16, but for the metallicity (Z) of galaxies. The anti-correlation of errors with the true values for this poorly constrained parameter is mainly caused by the limited parameter range (see text for more discussions about this).

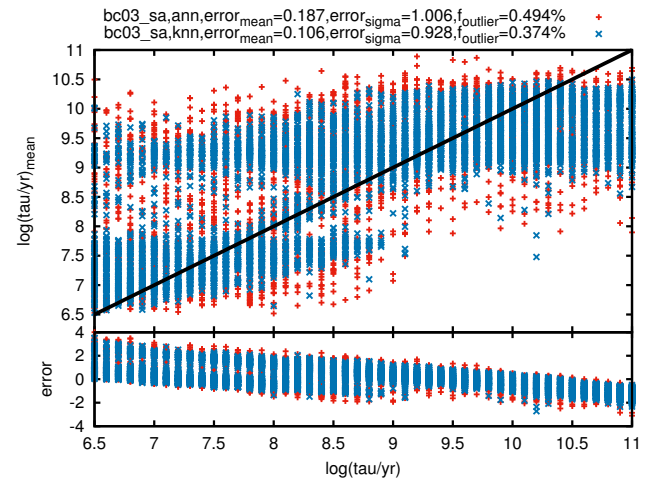


Figure 19. Similar to Figure 16, but for the e-folding time (τ) of the star formation history of galaxies.

ability of BayeSED to recover the free and derived parameters of a stellar population synthesis model from its observed multi-wavelength photometric SEDs by em-

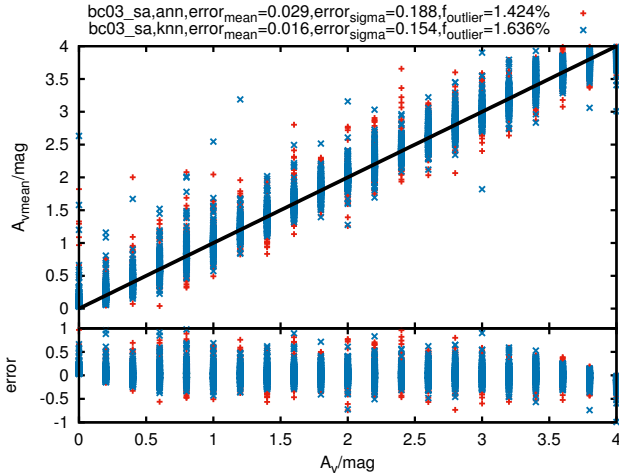


Figure 20. Similar to Figure 16, but for the dust extinction (A_v) of galaxies.

ploy mock sample of galaxies. Generally, we believe that the results obtained with BayeSED code are acceptable. Indeed, there are some extreme outliers in the recovered values, and they could be partly resulted from the BayeSED code itself. However, we believe that the errors in the recovered value of parameters are dominated by the nature of these parameters themselves and the limited information about them in the photometric data.

4. APPLICATION TO A KS-SELECTED SAMPLE IN THE COSMOS/ULTRAVISTA FIELD

With the systematic tests of BayeSED code in §3, we believe it should be able to obtain reliable results for real galaxies. In this section, we will apply BayeSED to interpret the SEDs of galaxies in a Ks-selected sample in the COSMOS/UltraVISTA field. Since the tests in §3 showed that the results obtained with KNN method have larger Bayesian evidence than that obtained with ANN, we only show the results obtained with KNN method in this section.

4.1. A Ks-selected Catalog in the COSMOS/UltraVISTA Field

The UltraVISTA survey (McCracken et al. 2012) is a NIR sky survey with a unique combination of area and depth. When fully complete, the survey will cover an area of 1.8 deg^2 down to $K_s \sim 24.0$. Meanwhile, the survey field of UltraVISTA is the COSMOS field (Scoville et al. 2007), which has the most extensive multi-wavelength coverage and is an attractive field for the studies of distant galaxies. Muzzin et al. (2013) presented a catalog covering 1.62 degree^2 of the COSMOS/UltraVISTA field. The catalog provides photometry in 30 photometric bands including the available GALEX, Subaru, Canada-France-Hawaii Telescope, VISTA, and Spitzer data. The sources in the catalog have been selected from the DR1 UltraVISTA K_s band imaging that reaches a depth of $K_{s,tot} = 23.4 \text{ AB}$ with 90% completeness.

In the study of this section, we have selected a subset of Muzzin et al. (2013) catalog with `star` = 0, `contamination` = 0, `nan_contam` < 5, $K_s < 23.9$ (5σ depth of the survey). These objects are considered to be galaxies with good photometry. Besides, we only

selected objects with known spectroscopic redshifts for this illustrative study. This results in a sample of 5467 galaxies with $0 < z < 2$. A more comprehensive study of all galaxies in the Muzzin et al. (2013) catalog with BayeSED will be presented in a future work. In addition to the photometric catalog, Muzzin et al. (2013) also provide a catalog of photometric redshifts computed with the EAZY code (Brammer et al. 2008) and a catalog of stellar masses and stellar population parameters determined using the FAST SED-fitting code (Kriek et al. 2009) for all galaxies in the survey. We will compare our results obtained with BayeSED code with that of Muzzin et al. (2013) obtained with FAST code.

4.2. Interpreting and Results

To be compared with the results of Muzzin et al. (2013), we have used the Bruzual & Charlot (2003) or Maraston (2005) SED models with solar metallicity, Calzetti et al. (2000) dust extinction law, and an exponentially declining star formation history to interpret the SEDs of galaxies in our sample. Chabrier (2003) or Salpeter (1955) IMF has been used for the bc03 model, while Kroupa (2001) or Salpeter (1955) IMF has been used for ma05 model. In total, four free parameters are involved, including $\log(\tau/\text{yr})$ in the range of [6.5, 11], $\log(\text{age}/\text{yr})$ in the range of [7.0, 10.1], and visual attenuation A_v in the range of [0, 4]. During the sampling with MultiNest, the four parameters are allowed to be uniformly and continuously selected from the allowed parameter space. As another prior, the age of a galaxy is forced to be less than the age of the universe at the redshift of the galaxy. In BayeSED, the scale factors of model SEDs are not considered as free parameters during the sampling of parameter space with MultiNest. Instead, they are uniquely determined using the efficient iterative algorithm of Sha et al. (2007) as in the EAZY code. With this algorithm, a linear combination of multiple SED models can be used to interpret the observed SEDs of galaxies. In this paper, only the stellar population synthesis model of bc03 or ma05 model is employed to interpret the observed SEDs. So, to be consistent with this selection of SED model, the Spitzer bands with wavelengths longer than $4.5 \mu\text{m}$ have not been used during the fitting of SEDs, since no model for dust emission has been used.

4.2.1. Comparison with the results of Muzzin et al. (2013)

In §3, we have verified the internal consistency of BayeSED by using a mock sample of galaxies. Here, we instead check the external consistency of BayeSED with the widely used FAST code as employed by Muzzin et al. (2013). Due to the very different methodologies employed by the two codes, we expect some differences for the results obtained by them. However, the results obtained by the two codes should be generally consistent with each other.

In Figures 27, 28, 29, 30, and 31, we compare the values of age, e-folding time, dust extinction, stellar mass, and star formation rate obtained using BayeSED with that obtained using FAST by Muzzin et al. (2013). Only the results obtained with Bruzual & Charlot (2003) model and Chabrier (2003) IMF have been shown. As shown in the figures, except for a small fraction of extreme outliers,

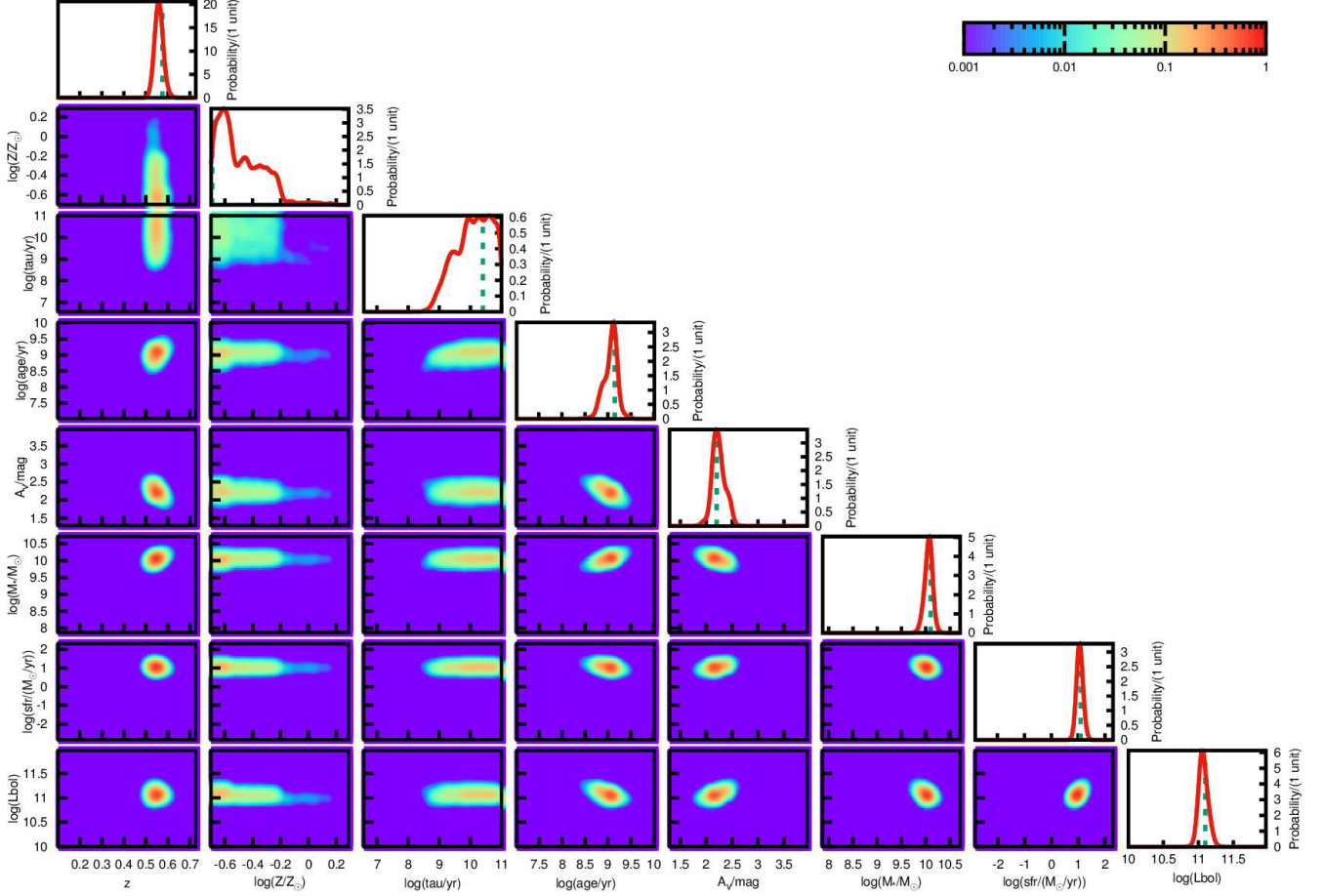


Figure 21. The posterior probability density functions (PDFs) of parameters for a mock galaxy with $z = 0.58$, $\log(Z/Z_\odot) = -0.70$, $\log(\tau/\text{yr}) = 10.4$, $\log(\text{age}/\text{yr}) = 9.15$, $A_V/\text{mag} = 2.2$, $\log(M_*/M_\odot) = 10.1$, $\log(\text{sfr}/(M_\odot/\text{yr})) = 1.1$, and $\log(L_{\text{bol}}/L_\odot) = 11.1$. Except for the metallicity (Z) and e-folding time (τ), the PDFs of all other parameters show a sharp peak around the true values. The PDFs of $\log(\text{age}/\text{yr})$ and A_V/mag show a weak second peak, indicative of the degeneracy between them. Meanwhile, the PDFs of $\log(Z/Z_\odot)$ and $\log(\tau/\text{yr})$ show even more peaks, indicative of serious degeneracies with other parameters.

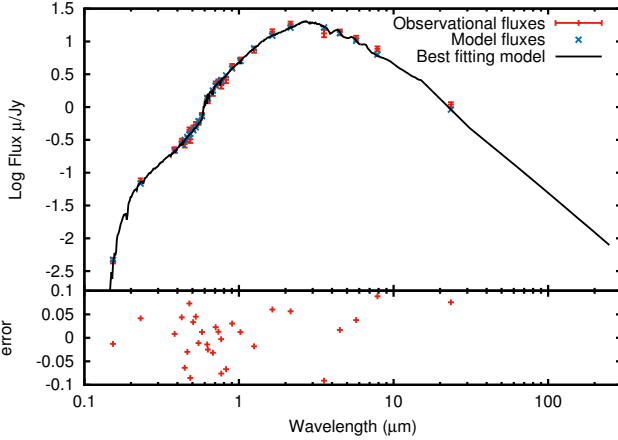


Figure 22. The best fitting results for the same mock galaxy as in Figure 21.

our results are generally consistent with that of Muzzini et al. (2013). This is more clear for those parameters that are more likely to be well constrained, such as stellar mass. Besides, the estimation of e-folding time (τ) seems very similar for the two codes. It is much better than the case for a mock sample of galaxies as shown in Figure 19. This is mainly because the redshift of a

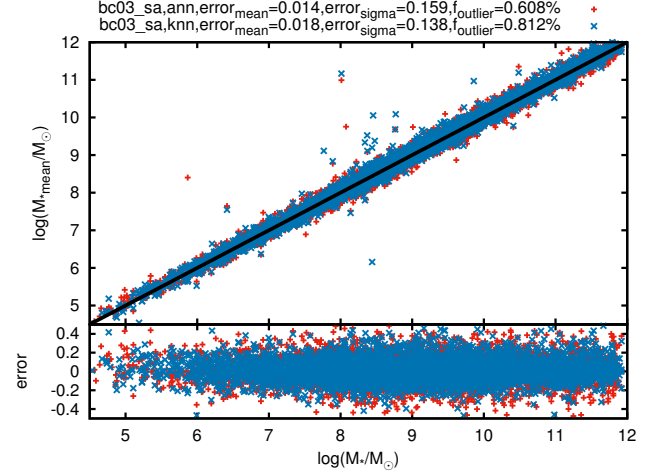


Figure 23. The derived stellar mass ($\text{mass}_{\text{mean}}$) compared with the true values (mass) of the mock sample of galaxies. The stellar mass is estimated with the mean of the posterior distribution of stellar mass. The bc03 model with a Salpeter (1955) IMF was used in this example. On the top of this figure, the mean, dispersion and the fraction of outliers with an error larger than 3σ of the distribution of errors are shown.

galaxy is fixed to the spectroscopic redshift, instead of as an additional free parameter that needs to be estimated

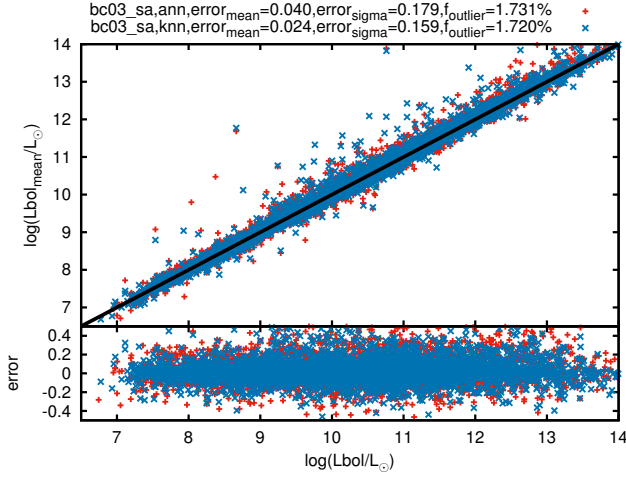


Figure 24. Similar to Figure 23, but for the bolometric luminosity (Lbol).

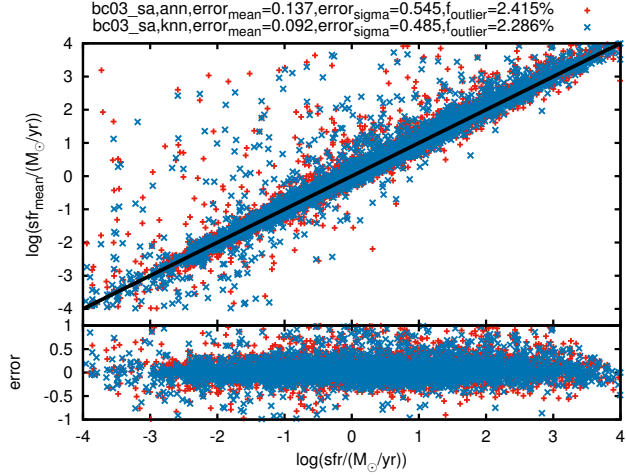


Figure 25. Similar to Figure 23, but for the star formation rate (sfr).

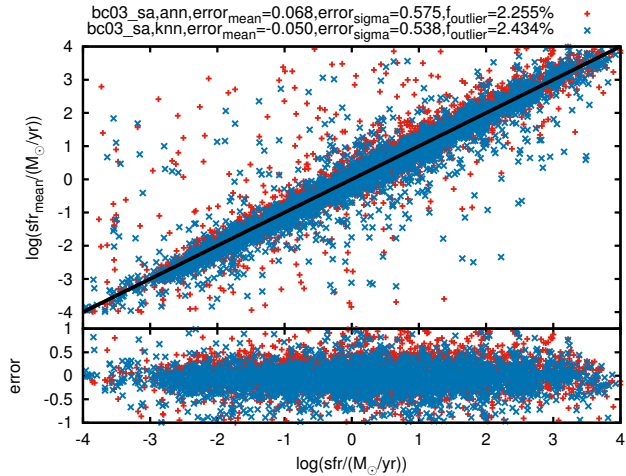


Figure 26. Similar to Figure 25, but now the MAL estimations of the star formation rates of galaxies have been used.

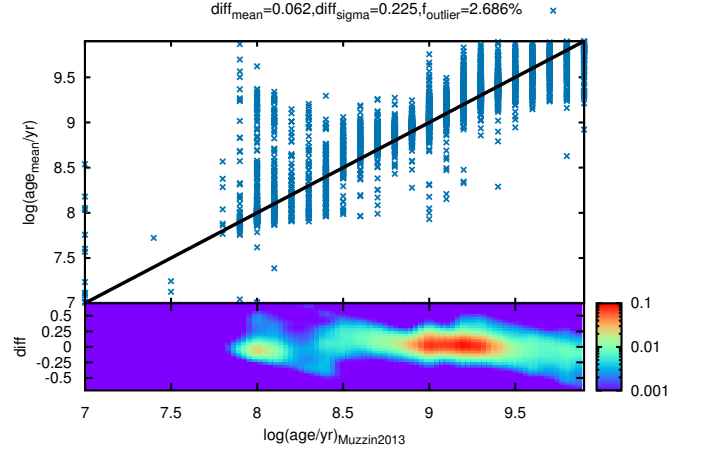


Figure 27. The age of galaxies obtained with our BayeSED code compared with that obtained with the FAST code by Muzzin et al. (2013). The parameter is estimated with the mean of the posterior distribution. The probability density distribution of the difference between the results of the two codes is shown in the lower panel. On the top of the figure, the mean, dispersion and the fraction of outliers with a difference larger than 3σ is shown.

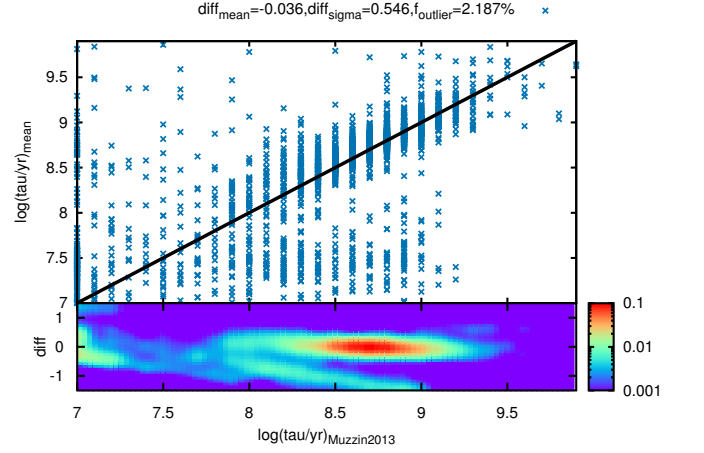


Figure 28. Similar to Figure 27, but for the e-folding time tau.

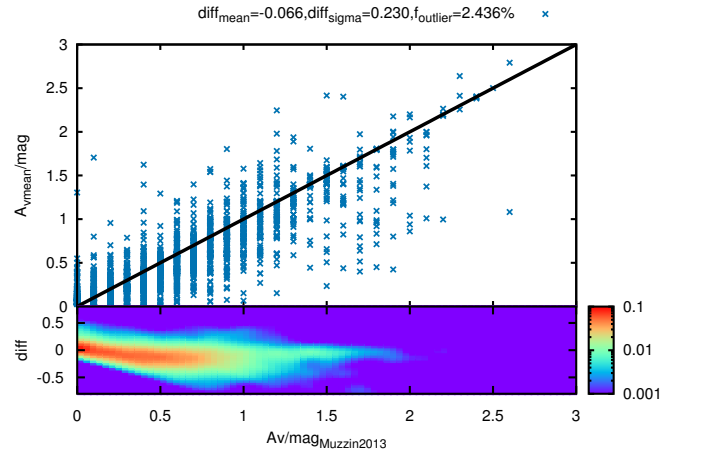


Figure 29. Similar to Figure 27, but for the dust extinction A_v .

from the observations.

An important advantage of BayeSED over traditional grid-based SED-fitting methods is that the parameter space of SED models can be sampled extensively and

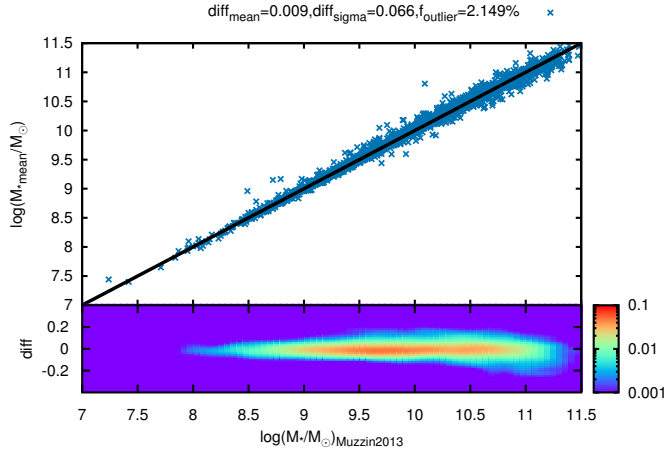


Figure 30. Similar to Figure 27, but for the stellar mass.

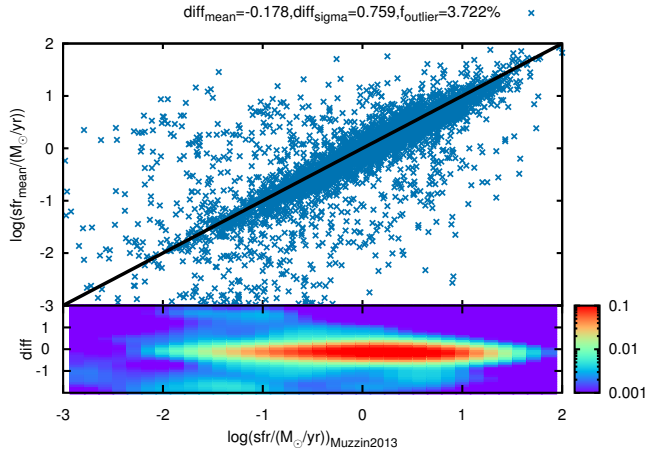


Figure 31. Similar to Figure 27, but for the star formation rate.

continuously. With the grid-base methods, the estimation of parameters are only allowed to be within the pre-computed set of grid points. This is a very unnatural restriction when interpreting the SEDs of real galaxies, and can result in biased results. In Figure 32, we have shown the distribution of stellar mass vs. star formation rate with results obtained with BayeSED and that obtained with grid-base FAST SED-fitting code. Generally, the two set of results are consistent with each other, especially for the star-forming main sequence. However, the results obtained with grid-base code show some unnatural parallel groups, which does not exist in results obtained with BayeSED. Actually, this kind of issues are even more clear for the free parameters of a SED model.

As mentioned in §1, there are many uncertainties in the population synthesis modeling of galaxy SEDs. So, with the ability of BayeSED code for an efficient computation of Bayesian evidence in hand, it would be very interesting to do a Bayesian model comparison for different population synthesis modeling. In Figure 33, we show the probability density distribution function (PDF) and cumulative distribution function (CDF) of Bayes factor $\ln(B)$ for 5467 galaxies in the Ks-selected sample with spectroscopic redshift. Four different combinations of population synthesis model and IMF are considered as four different models for Bayesian model comparison. The

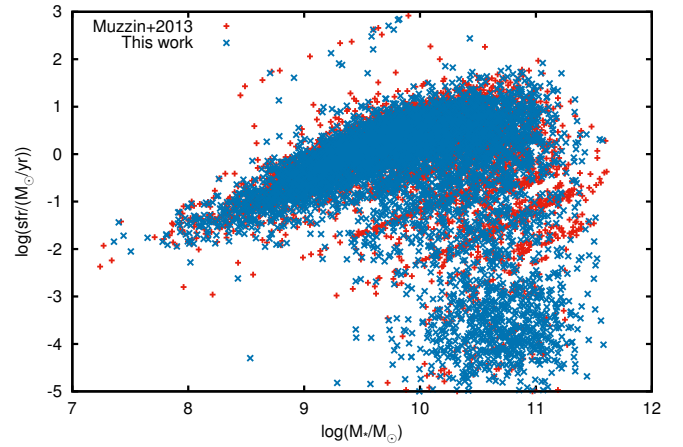


Figure 32. The distribution of galaxies in the M_* -SFR diagram. The results obtained with BayeSED in this work are generally consistent with those of Muzzin et al. (2013) who employed the FAST code, while the former show a more natural distribution.

Bruzual & Charlot (2003) model with Chabrier (2003) IMF and solar metallicity, the one used in the work of Muzzin et al. (2013), is used as the base model for the computation of all Bayes factors hereafter in this paper. The distributions show that the base model statistically has larger Bayesian evidence than the other three models, since only 45%, 40%, and 24% of galaxies support them, respectively. In general, the Bruzual & Charlot (2003) model statistically has larger Bayesian evidence than Maraston (2005) model, no matter which IMF has been used. The Maraston (2005) includes a more advanced treatment of the thermally pulsating AGB stars. However, whether it is a better treatment of this phase remains an open issue (Kriek et al. 2010; Zibetti et al. 2013). The results in Figure 33 show that Maraston (2005) model is better than Bruzual & Charlot (2003) model only for less than 10% of galaxies in sample.

4.2.2. Metallicity and redshift as additional free parameters

For many works of SED-fitting of galaxies, the solar metallicity of stellar population is commonly assumed, even for high redshift galaxies. Apparently, this is not a very reasonable assumption for galaxies in the real universe. An excuse for this assumption is that it is usually very hard to determine the stellar population metallicity of galaxies with photometric data only. On the other hand, the SED-fitting of galaxies would be much more time-consuming, especially for grid-base methods. However, biased results could be obtained if solar metallicity is assumed for all galaxies. To test the importance of the assumption about metallicity, we have employed BayeSED code to interpret the SEDs of the same sample of galaxies, but with metallicity as an additional free parameter ranging from $0.2Z_\odot$ to $2Z_\odot$. In practice, we found that this will not obviously increase the time of computation for our BayeSED code. In Figure 34, we show the PDF and CDF of Bayes factor $\ln(B)$ for the same sample of galaxies in this case. Compared with the results in Figure 33, the Bayesian evidences of all models are clearly increased. Now, the two bc03 models statistically have larger Bayesian evidence than the base model. Meanwhile, for the two ma05 models, only the one with Kroupa (2001) IMF has slightly larger support

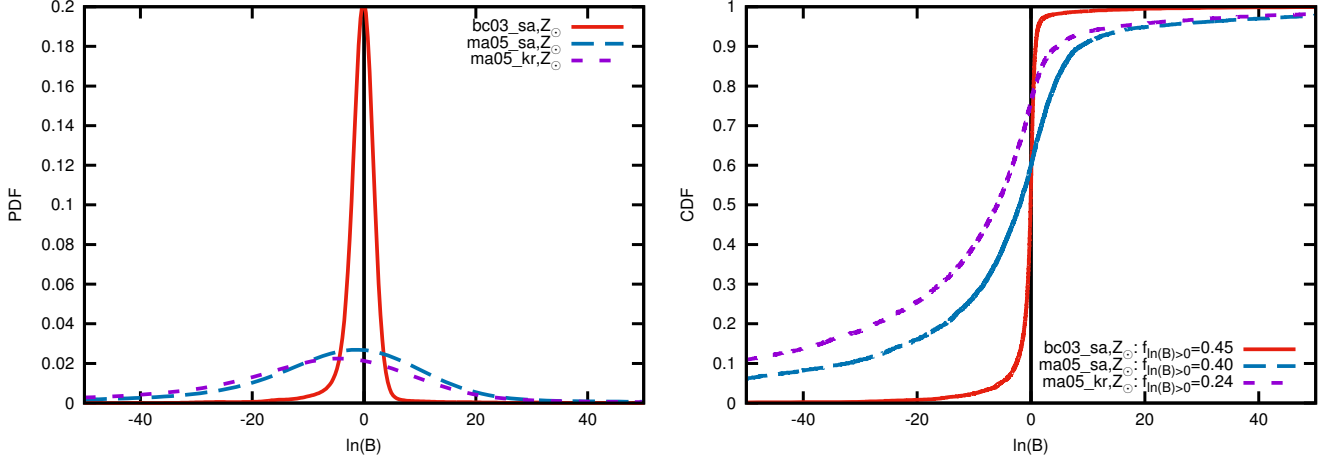


Figure 33. The probability density distribution function (PDF) and cumulative distribution function (CDF) of Bayes factors $\ln(B)$ for 5467 galaxies in the Ks-selected sample with spectroscopic redshift. The four combinations of SED model and IMF are considered as four different models for Bayesian model comparison. The [Bruzual & Charlot \(2003\)](#) model with a [Chabrier \(2003\)](#) IMF and solar metallicity is used as the base model for the computation of all Bayes factors hereafter in this paper. The distributions show that the base model statistically has larger Bayesian evidence than the other three models, since only 45%, 40%, and 24% of galaxies support them, respectively. In general, the [Bruzual & Charlot \(2003\)](#) model statistically has larger Bayesian evidence than the [Maraston \(2005\)](#) model, regardless of which IMF has been used.

rate than the base model. Besides, it is worth to notice that [Maraston \(2005\)](#) model seems more sensitive to the different selection of IMF than the [Bruzual & Charlot \(2003\)](#) model. Generally, with metallicity as an additional free parameter, it becomes more clear that the [Bruzual & Charlot \(2003\)](#) model has statistically higher Bayesian evidence than [Maraston \(2005\)](#) model.

As mentioned in §2.4, the redshift of galaxy can be set as a free parameter as well in BayeSED code. So, it is possible to simultaneously obtain photometric redshift and stellar population parameters of a galaxy with BayeSED, while using the same set of SED models and so more self-consistently. Here, we test the reliability of BayeSED for the determination of photometric redshift of galaxies. The distribution of Bayesian evidences for this case are shown in Figure 35. Compared with the results shown in Figure 34, where the spectroscopic redshifts have been used, the Bayesian evidence of all models decrease a little in this case. In Figure 36, we compared the estimated photometric redshifts with the spectroscopic redshifts for galaxies in our sample. The performance of a code for photometric redshift estimation is usually judged by the root mean square (RMS) of $(z_p - z_s)/(1 + z_s)$. In our case, $\sigma_{\text{RMS}} = 0.0449$. When outliers with error larger than 3σ ($48/5467=0.88\%$) are removed, $\sigma_{\text{RMS}} = 0.0254$. This outperforms the result of [Benítez \(2000\)](#) (0.06) who have employed a Bayesian approach as well and employed more informative physical priors. Our results is not as good as that of [Muzzin et al. \(2013\)](#) ($\sigma_{\text{RMS}} = 0.013$), who have employed the EAZY code and considered the effect of zero-point offset from an iterative procedure. However, they only considered 5119 galaxies with high-quality spectroscopic redshifts and uncontaminated photometry. A better judgement for the performance of photometric redshift estimation can be achieved by using the normalized median absolute deviation of $\Delta z = z_p - z_s$, which is defined ([Brammer et al. 2008](#)) as:

$$\sigma_{\text{NMAD}} = 1.48 \times \text{median} \left(\left| \frac{\Delta z - \text{median}(\Delta z)}{1 + z_p} \right| \right). \quad (10)$$

σ_{NMAD} is less sensitive than σ_{RMS} to the outliers. In our case, $\sigma_{\text{NMAD}} = 0.0185$, while the median of the error is -0.0135 and the fraction of outliers is 1.45%. Except for the larger value of median, this is better than the result of [Brammer et al. \(2008\)](#) with EAZY code, which is applied to a smaller sample with 1989 galaxies but spans a wider range of redshift. Given the difficulty of using a population synthesis model instead of using carefully selected templates or employing some kind of empirical training procedure to estimate the photometric redshift of galaxies, the degree of accuracy achieved by BayeSED is acceptable. In the future version of BayeSED, we plan to add more optimizations that have been adopted by many codes for a better estimation of photometric redshift.

4.2.3. The distributions of galaxy color and stellar population parameters

The observed distributions of galaxy properties provide important clues for understanding the formation and evolution of galaxies, and benchmarks for discriminations between different semi-analytic modelings and hydrodynamical simulations of galaxy formation. However, a detailed mathematical characterization of the distributions of these properties and a full theoretical understanding of them in the context of galaxy formation and evolution ([Somerville et al. 2008](#); [Schaye et al. 2010](#); [Buzzoni 2011](#)) is beyond the scope of this paper. Here, we show some well known features for the distribution of galaxy properties that are obtained with BayeSED code. For the results presented in this subsection, the spectroscopic redshifts of galaxies have been used, and the metallicity is set to be a free parameter ranging from $0.2Z_{\odot}$ to $2Z_{\odot}$. Besides, only the results obtained with [Bruzual & Charlot \(2003\)](#) model and [Chabrier \(2003\)](#) IMF have been presented, since it has the largest Bayesian evidence as shown in Figure 34.

In Figure 37, we show the distribution of the rest-frame $U - B$ color of galaxies in our sample. The well known bimodal distribution of the color of galaxies is clearly shown in the figure. These galaxies can be divided to

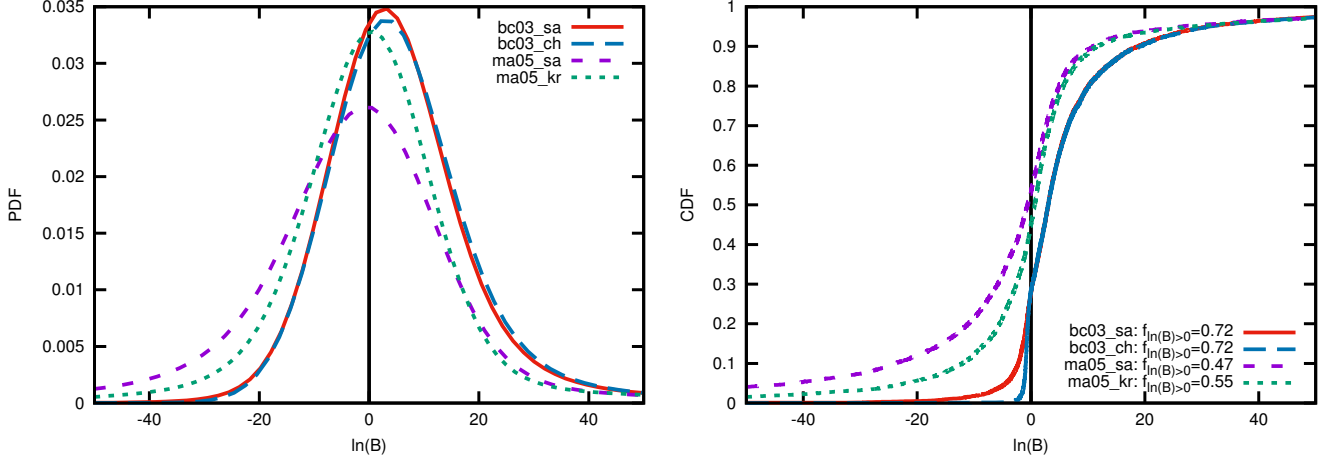


Figure 34. Similar to Figure 33, but now the metallicity is set to be an additional free parameter ranging from $0.2Z_{\odot}$ to $2Z_{\odot}$. It is clear that the Bayesian evidence of all models have increased significantly. Now, the two bc03 models have a much higher support rate than the base model. Meanwhile, for the two ma05 models, only the one with a Kroupa (2001) IMF has a slightly higher support rate than the base model. So, it becomes even clearer that the Bruzual & Charlot (2003) model statistically has a larger Bayesian evidence than the Maraston (2005) model. Meanwhile, it seems that the latter model is more sensitive to the different choice of IMF than the former model.

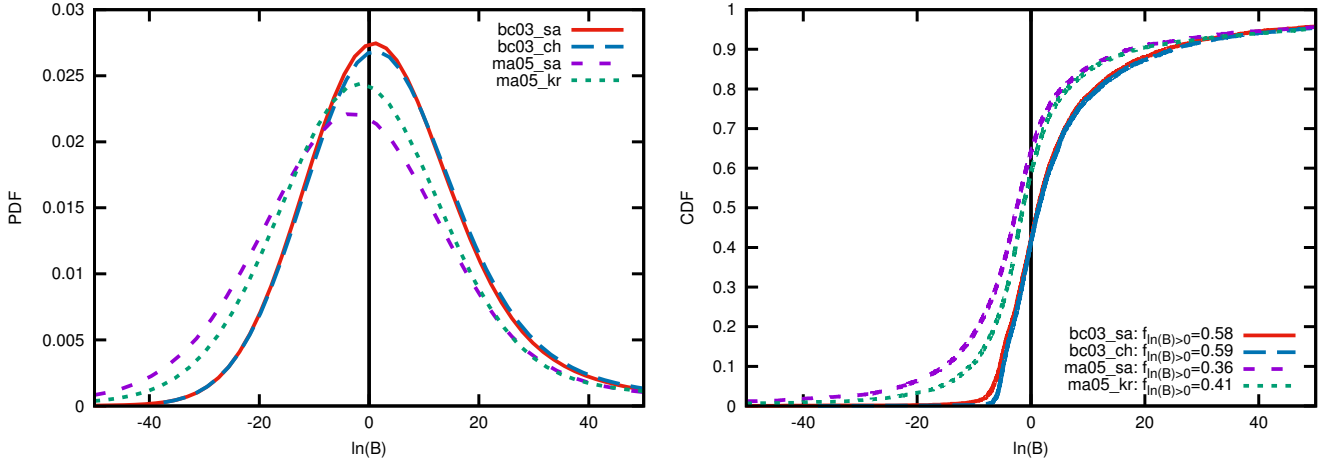


Figure 35. Similar to Figure 34, but now the redshift is also set to be an additional free parameter. In this case, the Bayesian evidence of all models have decreased a little. This demonstrates that an additional free parameter not necessarily increases the Bayesian evidence of a model if the increased complexity of the model is not rewarded with a much better fitting to the observations.

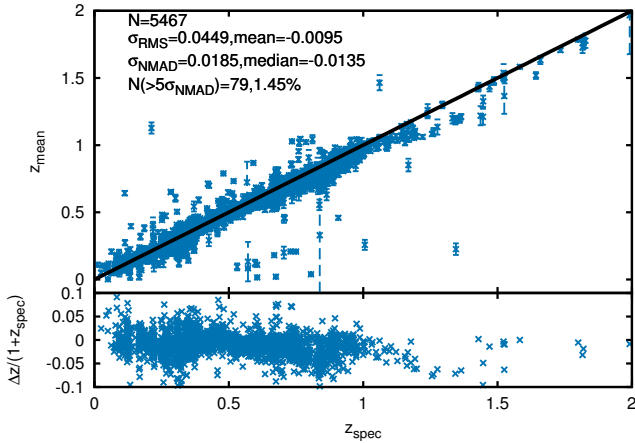


Figure 36. Photometric redshifts vs. Spectroscopic redshifts for galaxies in our sample. By using BayeSED, it is now possible to obtain photometric redshift and the stellar population parameters all simultaneously and self-consistently. The performance of BayeSED for the estimation of photometric redshift is shown in the figure (see text for more details).

the red ones with $U - B \gtrsim 1.1$, and the blue ones with $U - B \lesssim 1.1$. The red galaxies and blue are thought to be fundamentally different populations of galaxies. This can be more clearly noticed in the M_* -SFR diagram, as shown in Figure 38. Generally, blue galaxies are mainly the star-forming galaxies in the “main sequence”, while red galaxies are mainly the quiescent galaxies with negligible, if any, star formation. Meanwhile, there are a few galaxies with $SFR > 10$, above the “main sequence”. These galaxies have the bluest color in our sample, and should be the starburst galaxies. On the other hand, a portion of red galaxies are actually star-forming galaxies with $M_* \gtrsim 3 \times 10^{10} M_{\odot}$.

It is well known that many physical parameters, such as metallicity, dust attenuation, and age, can affect the color of galaxies. These parameters are degenerated with each other to a certain extent, preventing them to be determined accurately all simultaneously. However, they should be accurate enough to do some qualitative analysis. Moreover, we can check if the distributions of these parameters are reasonable in the context of galaxy evo-

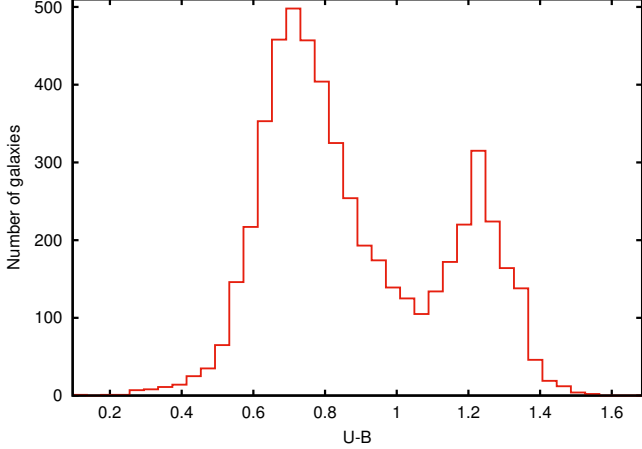


Figure 37. The bimodal distribution of rest-frame $U - B$ colors for galaxies in our sample.

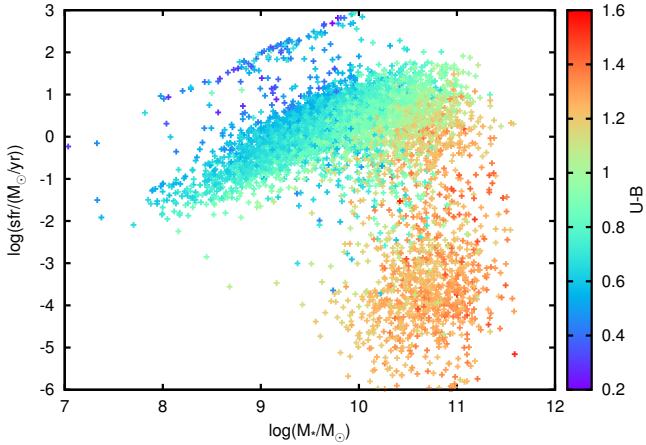


Figure 38. The distribution of rest-frame $U - B$ colors in the M_* -SFR diagram.

lution.

As shown in Figure 18, it is usually hard to determine the stellar metallicity of galaxies with photometric data only. However, since the spectroscopic redshift of galaxies have been used, we expect that the estimated metallicity of galaxies here should be more accurate. In Figure 39(a), we show the distribution of stellar metallicity of galaxies in the M_* -SFR diagram. As clearly shown in the figure, the blue galaxies all have low metallicity, while the red galaxies have much higher, near solar, metallicity. This is reasonable in the context of galaxy evolution. Since the blue galaxies are mainly star-forming and starburst galaxies, the stars in these galaxies should be formed recently from low-metallicity gas. On the other hand, red galaxies are fully evolved galaxies, where many stars are formed from recycled-gas enriched by the last generations of star formation. Meanwhile, in Figure 39(b), we show the distribution of dust extinction of the same galaxies in the M_* -SFR diagram. The starburst galaxies have the highest dust extinction, while the more massive star-forming galaxies show less but still important dust extinction. In contrary, there are very little dust extinction in both the quiescent and low mass star-forming galaxies.

Usually, the stellar population age of galaxies is thought to be the main parameter for the color of galax-

ies, since galaxies become redder when they aging. However, as shown in Figure 39(c), it seems not easy to separate galaxies into red and blue with age alone. The starburst galaxies, which are the youngest in our sample, are indeed blue galaxies. However, some blue low mass star-forming galaxies, instead of the reddest quiescent galaxies, are the oldest galaxies in our sample. As related to age, in Figure 39(d), we show the distribution of the e-folding time of SFH in the M_* -SFR diagram. Interestingly, the galaxies in our sample can be clearly divided into at least five groups with this distribution. The blue star-forming galaxies in the main sequence (G1) have the longest e-folding time of SFH. On the other hand, the red quiescent galaxies (G2) have much shorter e-folding time of SFH. Most galaxies in our sample belong to this two groups. The galaxies with the shortest e-folding time of SFH (G3) are located between the last two groups. These galaxies are also unique in the distribution of age, as shown in Figure 39(c). Meanwhile, the galaxies at the high-mass end of the main sequence (G4) constitute the four group. Finally, those starburst galaxies with very young age, highest dust extinction and much shorter e-folding time than normal star-forming galaxies are considered to be the fifth group of galaxies (G5).

The galaxies in the G3 and G4 groups are likely in the transition stages between red and blue galaxies. To check if they are related to AGN activities, we show the distribution of the apparent magnitude at $24\mu\text{m}$ in Figure 39(e). Meanwhile, we show the distribution of redshifts of these galaxies in Figure 39(f). For most of these galaxies, the fluxes at $24\mu\text{m}$ correspond to the rest-frame mid-IR fluxes, which are thought to be responsible by dust heated by AGN (Nenkova et al. 2002; Elitzur 2006; Fritz et al. 2006; Nenkov et al. 2008a,b; Han et al. 2012). As shown clearly in Figure 39(e), galaxies in the high-mass end of the star-forming main sequence have the strongest mid-IR emissions. These galaxies should be the composite galaxies. Most galaxies in the G4 group show important mid-IR emissions, while most galaxies in the G3 group do not. The galaxies in the G3 group could be galaxies at an earlier evolution stage, when the AGN activities are still weak or not launched yet.

5. SUMMARY AND CONCLUSIONS

In this paper, we have described an updated version of BayeSED. On the basis of a previous work (Han & Han 2012, 2013), we have presented a major update to this code in this paper. Firstly, the most up-to-date version of the MultiNest Bayesian inference (§2.1) tool has been employed. The MultiNest algorithm (§2.2) has recently been improved by the importance nested sampling to allow a more efficient sampling of high-dimensional parameter space and more accurate calculation of Bayesian evidence. Secondly, besides the ANN method, we have added the KNN method as another method for the interpolation of model SED libraries in BayeSED (§2.3). For both ANN and KNN, we have defined a file format to store all necessary information about a SED model such that almost any SED model can be easily used by BayeSED to interpret the observed SEDs of galaxies. Thirdly, the redshift of a galaxy can be set as a free parameter, and the effects of IGM and Galactic extinction have been considered. So, it is now possible to obtain the redshift and other physical parameters of a galaxy,

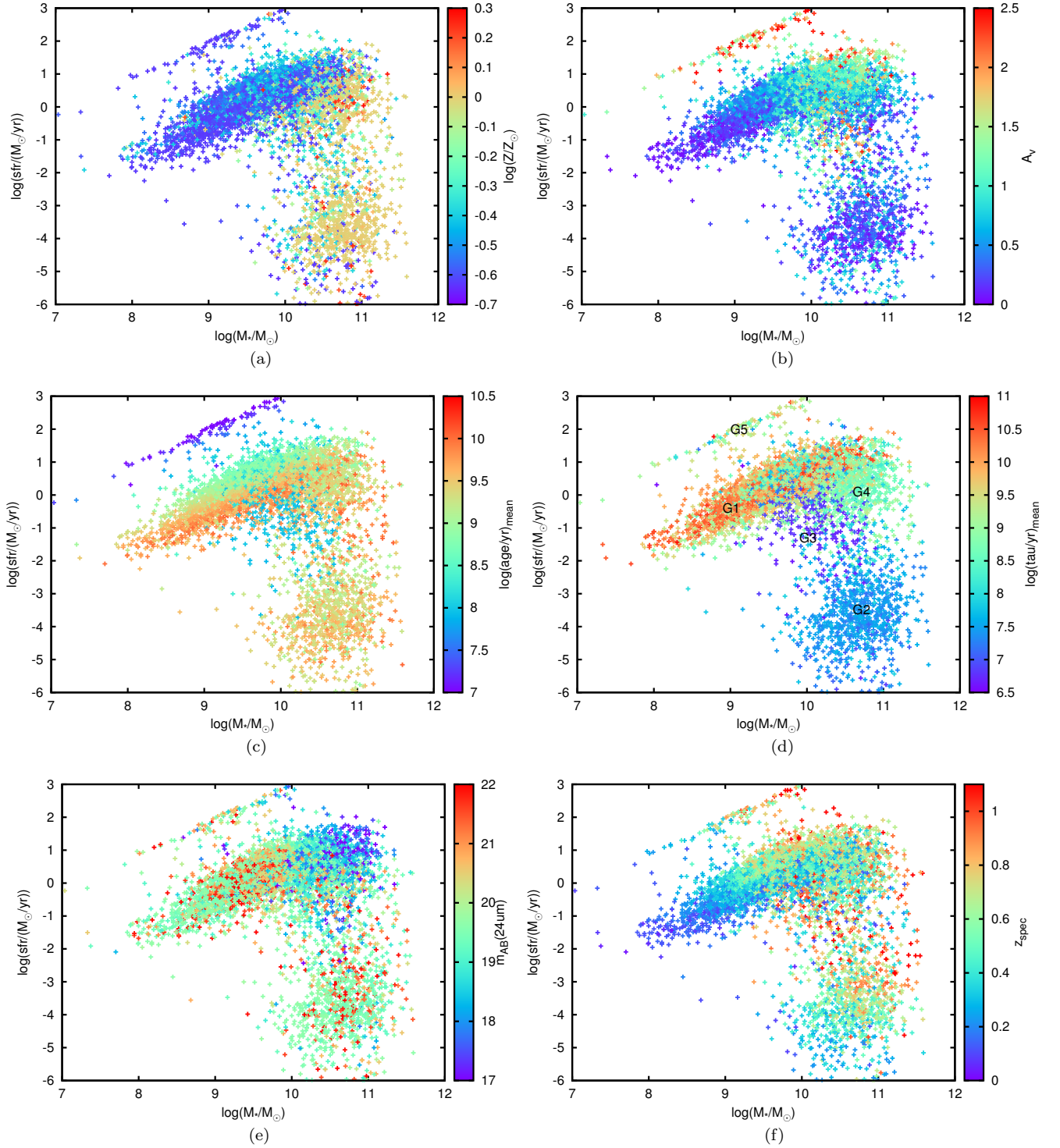


Figure 39. The distribution of age (a), metallicity (b), dust extinction (c), e-folding time of SFH (d), apparent magnitude at $24\mu\text{m}$ (e), and redshift (f) in the M_* -SFR diagram. With these distributions, five populations of galaxies can be easily recognized, which are more obvious in (d).

all simultaneously and self-consistently. Fourthly, the main body of BayeSED has been completely rewritten in C++ in an object-oriented programming fashion, and parallelized with MPI. So, with BayeSED, it is now possible to analyze the SEDs of a large sample of galaxies with a detailed Bayesian approach that is based on an intensive sampling of the parameter space of SED models.

We have systematically tested the reliability of the BayeSED code to recover the physical parameters of galaxies from their multi-wavelength photometric SEDs with a mock sample of galaxies (§3). For both the ANN and KNN methods, the tests show that BayeSED can recover the redshift and stellar population parameters reasonably well. We found that different parameters can be recovered with a varying degree of accuracy. While the redshift, age, dust extinction, stellar mass, bolometric luminosity and star formation rate can be properly recovered, it is usually hard to recover metallicity and e-folding time of SFH with photometric data, especially when the spectroscopic redshift is not available. We believe that this is due to the nature of these parameters themselves, and the limited information about them, in the photometric data. Meanwhile, the tests showed that the KNN interpolation method, although more memory- and time-consuming than the ANN method, may lead to more accurate results.

We systematically applied BayeSED to interpret the observed SEDs of a large sample of galaxies in the COSMOS/UltraVISTA field, with different evolutionary population synthesis models (§4). A Bayesian model comparison of evolutionary population synthesis models has been accomplished for the first time. We found that the Bruzual & Charlot (2003) model statistically has larger Bayesian evidence than the Maraston (2005) model for the Ks-selected sample of galaxies with spectroscopic redshifts and mostly less than one. Besides, we found that the Maraston (2005) model is more sensitive to the different choice of IMF than the Bruzual & Charlot (2003) model. However, the conclusion is drawn in a statistical sense and could be different for samples of galaxies at higher redshift (Maraston et al. 2006, 2010; Henriques et al. 2011), which is worthy of further investigation. Meanwhile, we found that, by using stellar metallicity as an additional free parameter, the Bayesian evidences of stellar population synthesis models can be increased significantly. Therefore, we conclude that it is important to set metallicity as a free parameter to obtain unbiased results, even if this parameter cannot be estimated very accurately with photometric data only.

We have compared our results obtained using our BayeSED code with that obtained using the widely used FAST code, and found a good agreement. Nevertheless, we found that the parameters estimated with BayeSED show more natural distributions than more conventional grid-based SED-fitting codes such as FAST. Besides, based on the rest-frame color and stellar population parameters obtained with BayeSED, we recognized five distinct populations of galaxies in the Ks-selected sample of galaxies. They may represent galaxies at different evolution stages or in different environments.

With the systematic tests using a mock sample of galaxies and the comparison with a popular grid-based SED-fitting code for a real sample of galaxies, we con-

clude that the BayeSED code can be reliably applied to interpret the SEDs of large samples of galaxies. Based on the MultiNest algorithm to allow intensive sampling of parameter space and efficient computation of Bayesian evidence of the SED models, BayeSED could be a powerful tool for investigating the formation and evolution of galaxies from the rich multi-wavelength observations currently available. Particularly, with the efficient computation of Bayesian evidence for SED models, BayeSED could be useful for the further development of evolutionary population synthesis models and other SED models for galaxies. Besides, while we have only applied BayeSED to the photometric data so far, it should be straightforward to apply similar methods to the spectroscopic data in the future.

We thank an anonymous referee for his/her valuable comments which help to improve the paper. The authors gratefully acknowledge the computing time granted by the Yunnan Observatories, and provided on the facilities at the Yunnan Observatories Supercomputing Platform. We thank Prof. Houjun Mo for helpful discussion about Bayesian evidence and Lulu Fan for helpful discussion about Bayesian SED-fitting of galaxies. This work is supported by the National Natural Science Foundation of China (Grant Nos. 11303084, U1331117, 11390374), the Science and Technology Innovation Talent Programme of the Yunnan Province (Grant No. 2013HA005), and the Chinese Academy of Sciences (Grant No. XDB09010202).

REFERENCES

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, **182**, 543
- Acquaviva, V., Gawiser, E., & Guaita, L. 2011, *ApJ*, **737**, 47
- Almeida, C., Baugh, C. M., Lacey, C. G., et al. 2010, *MNRAS*, **402**, 544
- Atwood, W. B., Abdo, A. A., Ackermann, M., et al. 2009, *ApJ*, **697**, 1071
- Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, *ApJ*, **600**, 681
- Baugh, C. M. 2006, *Reports on Progress in Physics*, **69**, 3101
- Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, *ApJS*, **149**, 289
- Benítez, N. 2000, *ApJ*, **536**, 571
- Bolzoni, M., Miralles, J.-M., & Pelló, R. 2000, *A&A*, **363**, 476
- Bonoli, S., Marulli, F., Springel, V., et al. 2009, *MNRAS*, **396**, 423
- Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, *MNRAS*, **370**, 645
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503
- Bruzual, A. G. 1983, *ApJ*, **273**, 105
- Bruzual, A. G. 2007, *Proceedings of the International Astronomical Union*, **2**, 125
- Bruzual, A. G., & Charlot, S. 1993, *ApJ*, **405**, 538
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Buzzoni, A. 1989, *ApJS*, **71**, 817
- . 1993, *A&A*, **275**, 433
- . 2011, *MNRAS*, **415**, 1155
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682
- Carballo, R., González-Serrano, J. I., Benn, C. R., & Jiménez-Luján, F. 2008, *MNRAS*, **391**, 369
- Catelan, M. 2009, *Ap&SS*, **320**, 261
- Cerviño, M. 2013, *NewAR*, **57**, 123
- Chabrier, G. 2003, *PASP*, **115**, 763
- Chen, X., & Han, Z. 2009, *MNRAS*, **395**, 1822
- Chen, Y.-M., Kauffmann, G., Tremonti, C. A., et al. 2012, *MNRAS*, **421**, 314
- Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, *MNRAS*, **358**, 363
- Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, *MNRAS*, **319**, 168
- Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, *MNRAS*, **362**, 505

- Collister, A. A., & Lahav, O. 2004, *PASP*, **116**, 345
- Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, *AJ*, **115**, 1693
- Conroy, C. 2013, *ARA&A*, **51**, 393
- Conroy, C., & Gunn, J. E. 2010, *ApJ*, **712**, 833
- Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, **699**, 486
- Conroy, C., White, M., & Gunn, J. E. 2010, *ApJ*, **708**, 58
- Croton, D. J., Springel, V., White, S. D. M., et al. 2006, *MNRAS*, **365**, 11
- Cybenko, G. 1989, *Mathematics of Control, Signals, and Systems (MCCS)*, **2**, 303
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, **388**, 1595
- Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, *ApJ*, **660**, L1
- Di Matteo, T., Springel, V., & Hernquist, L. 2005, *Nature*, **433**, 604
- Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *Astronomy and Geophysics*, **50**, 050000
- Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, *ApJS*, **180**, 306
- Dutton, A. A., Macciò, A. V., Mendel, J. T., & Simard, L. 2013, *MNRAS*, **432**, 2496
- Elitzur, M. 2006, *NewAR*, **50**, 728
- Elliott, D. L. 1993, A Better Activation Function for Artificial Neural Networks, Tech. rep., Institute for Systems Research, University of Maryland
- Fahlman, S. E. 1988, in Proceedings of 1988 Connectionist Models Summer School (Los Altos CA.: Morgan Kaufmann), 38
- Feroz, F., Gair, J. R., Graff, P., Hobson, M. P., & Lasenby, A. 2010, *Classical and Quantum Gravity*, **27**, 075010
- Feroz, F., Gair, J. R., Hobson, M. P., & Porter, E. K. 2009a, *Classical and Quantum Gravity*, **26**, 215003
- Feroz, F., & Hobson, M. P. 2008, *MNRAS*, **384**, 449
- Feroz, F., Hobson, M. P., & Bridges, M. 2009b, *MNRAS*, **398**, 1601
- Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, ArXiv e-prints, [arXiv:1306.2144](https://arxiv.org/abs/1306.2144) [astro-ph.IM]
- Fioc, M., & Rocca-Volmerange, B. 1997, *A&A*, **326**, 950
- Firth, A. E., Lahav, O., & Somerville, R. S. 2003, *MNRAS*, **339**, 1195
- Fitzpatrick, E. L. 1999, *PASP*, **111**, 63
- Friedman, J. H., Bentley, J. L., & Finkel, R. A. 1977, *ACM Trans. Math. Softw.*, **3**, 209
- Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, *MNRAS*, **366**, 767
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, **362**, 41
- Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, *ApJ*, **600**, L93
- Graff, P., Feroz, F., Hobson, M. P., & Lasenby, A. 2012, *MNRAS*, **421**, 169
- Gregory, P. 2005, Bayesian Logical Data Analysis for the Physical Sciences (New York, NY, USA: Cambridge University Press)
- Han, Y., Dai, B., Wang, B., Zhang, F., & Han, Z. 2012, *MNRAS*, **423**, 464
- Han, Y., & Han, Z. 2012, *ApJ*, **749**, 123
- Han, Y., & Han, Z. 2013, in *IAU Symposium, Vol. 295, IAU Symposium*, ed. D. Thomas, A. Pasquali, & I. Ferreras, 312
- Han, Z., Podsiadlowski, P., & Lynas-Gray, A. E. 2007, *MNRAS*, **380**, 1098
- Hastings, W. K. 1970, *Biometrika*, **57**, 97
- Haykin, S. 1999, Neural Networks: A Comprehensive Foundation (Upper Saddle River, NJ: Prentice Hall), 2nd edition
- Henriques, B., Maraston, C., Monaco, P., et al. 2011, *MNRAS*, **415**, 3571
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, **208**, 19
- Hopkins, P. F., Cox, T. J., Kereš, D., & Hernquist, L. 2008a, *ApJS*, **175**, 390
- Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, *ApJS*, **163**, 1
- Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008b, *ApJS*, **175**, 356
- Igel, C., & Hsken, M. 2000, in Proceedings of the second International Symposium on Neural Computation
- Jansen, F., Lumb, D., Altieri, B., et al. 2001, *A&A*, **365**, L1
- Jeffreys, H. 1961, The Theory of Probability, 3rd edn., Oxford Classics Series (Oxford: Oxford University Press)
- Johnson, S. P., Wilson, G. W., Tang, Y., & Scott, K. S. 2013, *MNRAS*, **436**, 2535
- Kaipio, J., & Somersalo, E. 2004, Statistical and Computational Inverse Problems, Applied Mathematical Sciences (Springer)
- Karpenka, N. V., March, M. C., Feroz, F., & Hobson, M. P. 2013, *MNRAS*, **433**, 2693
- Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, *MNRAS*, **353**, 713
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, **341**, 33
- Kavanagh, B. J. 2014, *Phys. Rev. D*, **89**, 085026
- Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, *A&A*, **501**, 1269
- Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2008, *MNRAS*, **385**, 1998
- Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, *ApJ*, **700**, 221
- Kriek, M., Labbé, I., Conroy, C., et al. 2010, *ApJ*, **722**, L64
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Kurt, & Hornik. 1991, *Neural Networks*, **4**, 251
- Larson, R. B., & Tinsley, B. M. 1978, *ApJ*, **219**, 46
- Lei, Z.-X., Chen, X.-F., Zhang, F.-H., & Han, Z. 2013, *A&A*, **549**, A145
- Leitherer, C., & Heckman, T. M. 1995, *ApJS*, **96**, 9
- Lewis, A., & Bridle, S. 2002, *Phys. Rev. D*, **66**, 103511
- Li, Z., & Han, Z. 2008, *MNRAS*, **387**, 105
- Li, Z., Mao, C., Chen, L., Zhang, Q., & Li, M. 2013, *ApJ*, **776**, 37
- Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., et al. 2003, *PASP*, **115**, 897
- Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, *MNRAS*, **283**, 1388
- Maraston, C. 1998, *MNRAS*, **300**, 872
- . 2005, *MNRAS*, **362**, 799
- Maraston, C., Daddi, E., Renzini, A., et al. 2006, *ApJ*, **652**, 85
- Maraston, C., Pforr, J., Renzini, A., et al. 2010, *MNRAS*, **407**, 830
- Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, *ApJ*, **619**, L1
- Martin, J., Ringeval, C., & Trotta, R. 2011, *Phys. Rev. D*, **83**, 063524
- Martin, R., & Heinrich, B. 1993, in Proceedings of the International Conference on Neural Networks (San Francisco, CA.: IEEE Press), 586
- Marulli, F., Bonoli, S., Branchini, E., Moscardini, L., & Springel, V. 2008, *MNRAS*, **385**, 1846
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, **544**, A156
- Meiksin, A. 2006, *MNRAS*, **365**, 807
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *The Journal of Chemical Physics*, **21**, 1087
- Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2013, *MNRAS*, **435**, 87
- Mo, H., van den Bosch, F., & White, S. 2010, Galaxy Formation and Evolution (The Edinburgh Building, Cambridge CB2 8RU, UK: Cambridge University Press)
- Mukherjee, P., Parkinson, D., & Liddle, A. R. 2006, *ApJ*, **638**, L51
- Murphy, E. J., Chary, R.-R., Alexander, D. M., et al. 2009, *ApJ*, **698**, 1380
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, *ApJS*, **206**, 8
- Myers, A. T., Krumholz, M. R., Klein, R. I., & McKee, C. F. 2011, *ApJ*, **735**, 49
- Neškova, M., Ivezić, Ž., & Elitzur, M. 2002, *ApJ*, **570**, L9
- Neškova, M., Sirocky, M. M., Ivezić, Ž., & Elitzur, M. 2008a, *ApJ*, **685**, 147
- Neškova, M., Sirocky, M. M., Nikutta, R., Ivezić, Ž., & Elitzur, M. 2008b, *ApJ*, **685**, 160
- Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, *A&A*, **507**, 1793
- Ocvirk, P., Pichon, C., Lançon, A., & Thiébaud, E. 2006, *MNRAS*, **365**, 46
- Padoan, P., Nordlund, A., & Jones, B. J. T. 1997, *MNRAS*, **288**, 145
- Pforr, J., Maraston, C., & Tonini, C. 2012, *MNRAS*, **422**, 3285
- . 2013, *MNRAS*, **435**, 1389
- Polletta, M., Tajer, M., Maraschi, L., et al. 2007, *ApJ*, **663**, 81
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, *Nature*, **323**, 533
- Salpeter, E. E. 1955, *ApJ*, **121**, 161
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, **402**, 1536
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Searle, L., Sargent, W. L. W., & Bagnuolo, W. G. 1973, *ApJ*, **179**, 427
- Serra, P., Amblard, A., Temi, P., et al. 2011, *ApJ*, **740**, 22
- Sha, F., Lin, Y., Saul, L. K., & Lee, D. D. 2007, *Neural Comput.*, **19**, 2004
- Shaw, J. R., Bridges, M., & Hobson, M. P. 2007, *MNRAS*, **378**, 1365
- Shen, S., Mo, H. J., White, S. D. M., et al. 2003, *MNRAS*, **343**, 978
- Silva, L., Schurer, A., Granato, G. L., et al. 2011, *MNRAS*, **410**, 2043
- Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, American Institute of Physics Conference Series, ed. U. V. T. R. Fischer, R. Preuss, 395

- Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, [MNRAS](#), **391**, 481
- Springel, V. 2010, [ARA&A](#), **48**, 391
- Springel, V., Di Matteo, T., & Hernquist, L. 2005a, [MNRAS](#), **361**, 776
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005b, [Nature](#), **435**, 629
- Tarantola, A. 2005, Inverse Problem Theory and Methods for Model Parameter Estimation (Society for Industrial and Applied Mathematics)
- Tian, B., Deng, L., Han, Z., & Zhang, X. B. 2006, [A&A](#), **455**, 247
- Tinsley, B. M. 1972, [A&A](#), **20**, 383
- Tinsley, B. M., & Gunn, J. E. 1976, [ApJ](#), **203**, 52
- Tojeiro, R., Heavens, A. F., Jimenez, R., & Panter, B. 2007, [MNRAS](#), **381**, 1252
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, [ApJ](#), **613**, 898
- Trotta, R. 2008, [Contemporary Physics](#), **49**, 71
- Trotta, R., Ruiz de Austri, R., & Pérez de los Heros, C. 2009, [JCAP](#), **8**, 34
- Verde, L., Peiris, H. V., Spergel, D. N., et al. 2003, [ApJS](#), **148**, 195
- Walcher, C. J., Böker, T., Charlot, S., et al. 2006, [ApJ](#), **649**, 692
- Walcher, C. J., Lamareille, F., Vergani, D., et al. 2008, [A&A](#), **491**, 713
- Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, [Ap&SS](#), **331**, 1
- Wild, V., Almaini, O., Cirasuolo, M., et al. 2014, [MNRAS](#), **440**, 1880
- Yèche, C., Petitjean, P., Rich, J., et al. 2010, [A&A](#), **523**, A14
- Zhang, F., Han, Z., Li, L., & Hurley, J. R. 2005, [MNRAS](#), **357**, 1088
- Zibetti, S., Gallazzi, A., Charlot, S., Pierini, D., & Pasquali, A. 2013, [MNRAS](#), **428**, 1479